

CAI
TB131
-81B01



PROGRAM EVALUATION

AN INTRODUCTION

BRANCH DISCUSSION PAPER



Office of the Comptroller
General of Canada

Bureau du Contrôleur
général du Canada

CAI
TB/31
- 8/301

Branch Discussion Paper 81-001

Table of Contents

1. WHAT IS PROGRAM EVALUATION?	1
2. DEVELOPMENT OF PROGRAM EVALUATION IN THE FEDERAL GOVERNMENT	2
3. BENEFITS FROM PROGRAM EVALUATION	3
4. THE CHALLENGE OF THE PROGRAM THEORY OF CHANGE/IMPACT	4
5. PLANNING PROGRAM EVALUATION STUDIES	5
5.1 The Nature of the Program to be Evaluated	5
5.2 The Questions to be Addressed	5
5.3 The Object of the Evaluation Study	5
5.4 The Evaluation Design	5
5.5 The Evaluation Process	5
5.6 The Role of the Evaluation Unit	5
5.7 Summary	5
6. LIMITATIONS AND ANALYTIC CHALLENGES FOR EVALUATION STUDIES	6
6.1 Program Objectives	6
6.2 Resources	6
6.3 Resource Constraints	6
6.4 Policy Implementation Challenges	6
6.5 Determining a Program's Worth	6
7. WHAT MAKES AN EVALUATION STUDY SUCCESSFUL?	7
CONCLUSIONS	8

Program Evaluation Branch
Office of the Comptroller General
February, 1981



Digitized by the Internet Archive
in 2024 with funding from
University of Toronto

PROGRAM EVALUATION: AN INTRODUCTION

Table of Contents

	<u>Page</u>
1. WHAT IS PROGRAM EVALUATION?	2
2. DEVELOPMENT OF PROGRAM EVALUATION IN THE FEDERAL GOVERNMENT	2
3. BENEFITS FROM PROGRAM EVALUATION	6
4. SOME EXAMPLES OF EVALUATION STUDIES OF GOVERNMENT PROGRAMS	7
5. PLANNING APPROPRIATE EVALUATION STUDIES	7
5.1 The Nature of the Program to be Evaluated	9
5.2 The Questions to be Addressed	10
5.3 The Client of the Evaluation Study	14
5.4 The Evaluation Method	15
5.5 The Resources and Time Available	16
5.6 The Need to Ensure Credibility	18
5.7 Summary	18
6. LIMITATIONS AND REALISTIC EXPECTATIONS FOR EVALUATION STUDIES	19
6.1 Program Objectives	20
6.2 Methodology	24
6.3 Resource Constraints	25
6.4 Using Evaluation Findings	26
6.5 Determining A Program's Worth	28
7. WHAT MAKES AN EVALUATION STUDY SUCCESSFUL?	29
FOOTNOTES	31

PROGRAM EVALUATION: AN INTRODUCTION

The purpose of this paper is to establish a common ground of understanding and a basic framework for discussing the subject of program evaluation. An earlier version of this paper was prepared as background material for consideration of several evaluation initiatives. This material can also serve as background material for two documents currently being prepared by the Office of the Comptroller General: "A Guide on the Program Evaluation Function in Federal Departments and Agencies" and "Principles for the Evaluation of Programs by Federal Departments and Agencies".

This discussion paper begins by describing what program evaluation entails (Section 1) and outlines the evolution of evaluation in the federal government (Section 2). It describes the benefits of program evaluation in Section 3 and gives examples of recent evaluation studies of government programs (Section 4). This is followed in Section 5 by a discussion of the main factors to be considered in planning an appropriate evaluation study. Section 6 discusses the limitation of conducting and using evaluation studies. The paper concludes (Section 7) with a list of characteristics of successful evaluation studies.

The paper discusses these topics, for the most part, in general terms and is not limited to describing program evaluation as it is carried out inside the federal government. Where appropriate, specific approaches to program evaluation are referred to, such as in the case of the Treasury Board policy on evaluating programs or evaluation studies that could be done for central agencies, Cabinet or Parliament.

The footnotes to this paper supply further information and references on this challenging and complex subject.

1. WHAT IS PROGRAM EVALUATION?

There is no agreement in the literature on what program evaluation should entail. The terminology used varies widely and is itself the cause of much confusion. Simply stated, program evaluation is the formal assessment of the continued relevance and of the effectiveness of existing programs. That is, it involves the systematic gathering of verifiable information on a program and demonstrable evidence on its results. The Office of the Comptroller General, in discussing government program evaluation, has expanded on this:¹

Program evaluation is the periodic, independent and objective review and assessment of a program to determine in light of present circumstance, the adequacy of its objectives, its design and its results both intended and unintended. Evaluations will call into question the very existence of the program. Matters such as the rationale for the program, its impact on the public, and its cost effectiveness as compared with alternative means of program delivery are reviewed.

In this context a program is a group of government activities, powers or functions that can be related in a logical fashion to a given objective or set of objectives. A (program) evaluation study is an analysis of specified aspects of a program and its effectiveness and an evaluation report is a report of the study, its findings and conclusions. Program evaluation, as a management function, is the evaluation of programs with respect to all relevant evaluation issues (see Table 2). Much of this paper focuses on evaluation studies rather than on the function of program evaluation. An expansion on these definitions is provided later (Section 5) after a number of additional concepts have been introduced.

2. DEVELOPMENT OF PROGRAM EVALUATION IN THE FEDERAL GOVERNMENT

The review and evaluation of existing programs has always been a part of managing in government. The nature and content of the reviews and evaluations have evolved from focusing mainly on the

resources used by programs (the dollars spent and numbers of people employed) to examining how the resources are used, the purposes of programs and their impacts and effects on society.

This change of emphasis has been necessary. The government has been called upon over the past three decades to provide an increasing array of goods and services to the public. Many current government programs reflect significant and sophisticated attempts to improve social and economic conditions in an increasingly complex society. As government activities have grown, the objectives and results of such programs have come under increased public scrutiny.

The increased concern with the purposes and hence the results of government expenditures can be traced to the 1960 Royal Commission on Government Organization (Glassco). Prior to Glassco, the analysis of programs relied mainly on the annual audits of government expenditures. A post-audit was performed for Parliament by the Auditor General and a pre-audit was conducted by the Comptroller of the Treasury. These audits ensured the probity of expenditures (the program inputs) but did not concentrate on the purposes and results of programs. The Glassco Commission's reports, published in 1962 and 1963, led to a general decentralization of financial and management control and to the development of more modern management techniques. The traditional concern for audit was coupled with a new concern about whether a program was conducted in a manner consistent with its legislated intent and specific operating goals.²

To assist in the reorientation of departmental managers to assume more responsibility for managing their affairs and to provide a better insight into the primary purposes of departmental operations, a new financial management approach - PPB, (Planning, Programming, Budgeting) - was introduced in 1968. "With its emphasis on defining program objectives and showing the full costs for each program"³, PPB

was intended to introduce planning and programming to the budgetary process. Its forward-looking multi-year basis was to provide a framework to facilitate the Treasury Board's allocation of resources.⁴ In practice, difficulties were encountered in trying to make the system operational and PPB did not live up to all its expectations.

However, PPB did contribute to the new emphasis on examining the intended results of programs. Since 1970-71, this has been reflected in the federal government's annual Estimates book which cross-classified expenditures by agency, by program, and by objects of expenditure.

During the 1970s there was an increase in evaluative activity throughout the government across the spectrum of program forecast and review, operational auditing, to in-depth, formal, quantitative evaluations of the effectiveness of specific programs. These latter formal evaluations were carried out on a highly selective basis normally in response to major policy planning priorities or issues rather than as part of the ongoing examination of existing programs by departmental management.

Thus while management in government has always carried out evaluation studies, formal, systematic, and periodic evaluation of programs was initiated in 1977 with the issuing of the Treasury Board Policy Circular 1977-47, "Evaluation of Programs by Departments and Agencies". To a considerable extent this policy was built upon the best practices already in existence in a number of departments and agencies. The general statement of the Policy is that:

Departments and agencies of the federal government will periodically review their programs to evaluate their effectiveness in meeting their objectives and the efficiency with which they are being administered.

The focus of the Policy is on the evaluation of programs, as opposed to, in particular, the evaluation of systems and procedures which are, in general, examined in internal audit.⁵ As envisaged in the policy, program evaluation is considered to cover a number of basic program evaluation issues:

- . Program Rationale: Does the program make sense?
- . Impacts and Effects: What has happened as a result of the program?
- . Objectives Achievement: Has the program performed as expected?
- . Alternatives: Are there better ways of achieving the results?

The purpose of program evaluation in the federal government undertaken pursuant to this policy is to ensure that deputy heads of departments and agencies have the appropriate information on the results of their programs in order to be able:

- . to account for the public monies for which they are responsible;
- . to make informed decisions on the management and resourcing of their programs; and
- . to provide quality advice to Ministers.

Changes in organizational structure have accompanied the evolution of the program evaluation function in the federal government. The Office of the Comptroller General was created in June 1978. This office has broad authority and responsibility for administrative practices and controls in the areas of financial management and procedures for program evaluation in the Government of Canada. More specifically, the Office has the responsibility to see that departments and agencies have established program evaluation procedures conforming to Treasury Board policy.

3. BENEFITS FROM PROGRAM EVALUATION

There is a growing acceptance that periodic program evaluation is a key function in the management of government departments. As has been observed⁶

Evaluation is necessary because of the scarcity of resources. ... evaluation involves the allocation of limited resources among competing activities.

The wider implementation of program evaluation will lead to a better understanding of the operations of programs and of the actual results of programs. This will contribute to:

- . improving accountability for government programs;
- . enhancing the ability of the government to allocate scarce resources for the greatest effects;
- . increasing the capability of the government to plan ahead and thereby to devise better programs to meet policy goals;
- . clarifying a program's contribution to policy initiatives; and
- . improving the management and administration of government programs, and thereby providing better government for Canadians.

These are the direct consequences. More and better evaluations will also have two more general consequences:

- . the regular involvement in the evaluation of programs will encourage a program's staff to examine critically their own assumptions and behaviour and thereby to experiment and to find ways of performing better; and
- . the acceptance of program evaluation by concerned citizens and groups as one legitimate and useful tool in the process of determining a program's worth to society, should help to reduce misunderstanding and disagreement over a program's effects.

4. SOME EXAMPLES OF EVALUATION STUDIES OF GOVERNMENT PROGRAMS

The diversity of types of evaluation studies carried out on federal programs is illustrated in Table 1. Evaluation studies are done by different people and for different reasons. For example, it is evident that none of the studies listed in Table 1 were undertaken pursuant to the 1977 Treasury Board policy, largely because of timing but also in several cases because the clients of the studies were not deputy heads. It is also clear that different evaluation studies use different techniques. The types of studies in the table range from determining the impacts on society to questioning the specific components or activities comprising the program. Evaluation techniques range from econometric modelling to interviewing beneficiaries of the program and experts in the subject-matter area.

This wide range of possible types of evaluation studies emphasizes the need for careful refinement of the term "program evaluation". It also implies that care must be taken to ensure that for any evaluation study an appropriate evaluation strategy is developed, one that will address the right issues and satisfy the client of the study.

5. PLANNING APPROPRIATE EVALUATION STUDIES

Planning for an evaluation study is perhaps the most critical phase of any evaluation. Evaluation studies can be costly in terms of both financial and human resources and experience has shown that many such studies in the past have turned out not to be useful (see Section 6.4). At least part of the reason for non-use of evaluation studies in the past has been a lack of adequate consideration before commencing the study as to what was needed, what could be done, and what would be done. Such planning to determine the proper focus and methodology - here called an evaluation assessment study - is essential for the design of an appropriate evaluation study.⁷

Table 1

SOME EVALUATION STUDIES OF GOVERNMENT PROGRAMS

Program Evaluated	Date	Client	Purpose of Evaluation	Evaluation Methodology
Department of Public Works Accommodation Program	1978	Senate Finance Committee	To describe the program's ability to meet its objectives and to assess the program's strengths and weaknesses	Interviews of experts and officials
Research and Development Incentive Grants Program	1976	Department of Industry, Trade and Commerce	To determine the effectiveness of R&D grants	Econometric analysis
Department of Fisheries & Environment (DFE)	1977/78	DFE and Treasury Board Secretariat	To identify from within its expenditure base the resources required to implement new initiatives	Analysis of department activities
Canada Manpower Training Program	1978	Employment and Immigration, Treasury Board Secretariat and Finance	To review the activities of the program in the context of the program's mandate	Sample surveys of participants
Unemployment Insurance Commission Benefits Program	1977	Economic Council	To determine the effect of UIC benefits on labour force participation	Econometric analysis

An appropriate evaluation study of a program requires consideration of:

- . the nature of the program to be evaluated;
- . the questions to be addressed;
- . the client of the evaluation study;
- . the evaluation method to be adopted;
- . the resources and time available for carrying out the study;
- and
- . the need to ensure credibility.

These six factors will assist those responsible for carrying out the evaluation study and the client to develop an appropriate evaluation strategy. The six factors are discussed in order below.

5.1 The Nature of the Program to be Evaluated

The first factor to be considered in designing an evaluation study is a definition or description of what is to be evaluated. What constitutes a "program" for evaluation purposes is not always obvious.⁸ A program may or may not coincide with the organizational structure of government or of a department. In particular, a program for evaluation purposes may not be a "program" as described in the government's annual Estimates. Within the federal government, considerable work has gone on to develop suitable program components for evaluation purposes pursuant to the Treasury Board policy on evaluation, i.e., groups of activities having a common objective established at the level of concern of the deputy head, and of appropriate size or impact to be a focus of decision making on programming at the departmental level. (In this discussion paper, the term program rather than program component is retained to describe what is evaluated.) The best disaggregation into program components suitable for evaluation purposes is often not obvious until after the department's activities are examined in some detail.

Having identified the activities which will constitute the program to be evaluated, it is necessary to have some understanding of how the program is supposed to work: an understanding of its basic rationale. That is, how the program activities produce program outputs which in turn result in particular impacts and effects on society. Objectives for the program can be specified in terms of the program outputs, the program's intended impacts and effects, or both, depending on the nature of the program and the level of objective being defined.

The state of development of the program must also be considered. Most programs cannot be evaluated in a meaningful way until they have been operating for some time.⁹ Furthermore, the type of service or good provided by the program may also influence the evaluation design,¹⁰ as will the type of intervention being used: expenditure, regulatory or tax expenditure.

5.2 The Questions to be Addressed

In evaluating a program a great number of questions could be asked and a wide array of information about the program's operations and results could be collected and analyzed. (As the examples in Table 1 illustrate, these could range from broad questions on the impacts of the program to more limited questions about the acquisition of resources for the program.) Basic to designing an appropriate study is a clear understanding of precisely which questions are to be addressed. This provides the necessary focus for the study.

Table 2, based on the four general evaluation issues mentioned earlier in Section 2, lists seven basic evaluation questions which can be addressed in a program evaluation. The table in fact provides an operational definition of the term "program evaluation", as it is being used in the federal government.

Table 2

BASIC PROGRAM EVALUATION ISSUES

Classes of Evaluation Issues	Basic Evaluation Questions
Program Rationale	<ul style="list-style-type: none">. To what extent are the objectives and mandate of the program still relevant?. Are the activities and outputs of the program consistent with its mandate and plausibly linked to the attainment of the objectives and the intended impacts and effects?
Impacts and Effects	<ul style="list-style-type: none">. What impacts and effects, both intended and unintended, resulted from carrying out the program?. In what manner and to what extent does the program complement, duplicate, overlap or work at cross-purpose with other programs?
Objectives Achievement	<ul style="list-style-type: none">. In what manner and to what extent were appropriate program objectives achieved as a result of the program?
Alternatives	<ul style="list-style-type: none">. Are there more cost-effective alternative programs which might achieve the objectives and intended impacts and effects?. Are there more cost-effective ways of delivering the existing program?

Program rationale issues focus on the continued relevance of the program in light of present social and economic conditions and government policy. Here the very existence of the program is to be questioned by asking if the program is still needed, even assuming it

is producing as expected. A good understanding of the rationale of the program should be developed, by comparing the current program activities with the mandated activities and examining the continued plausibility of the link between the program's outputs and both its objectives and other intended impacts and effects.

Consideration of the impacts and effects of a program implies a broad focus, an attempt to determine what has happened as a result of the program. It is concerned with the immediate impacts of the program and its more widespread, diffuse, or longer term effects, both intended and - often of more interest - unintended, regardless of the stated or claimed objectives of the program. This includes distributional conditions, as well as impacts and effects on and by related programs.

Objectives achievement issues are concerned with determining the manner and the extent to which appropriate objectives are achieved as a result of the program. This focus has often been narrowly taken as the only focus for program evaluation.

Finally, the consideration of alternatives is a yardstick for assessing the relative worth of the program. The objectives may have been met and there be no negative impacts or effects, but there may be better ways of achieving the objectives or intended impacts and effects: achieving the same results for less cost, obtaining more or better results for the same cost or achieving proportionately more or better results with increased costs. The program would then be more effective. Alternatives could include, as appropriate, both alternative ways of delivering the program and alternative programs to achieve the objectives and intended impacts and effects. These issues can be approached by considering questions of economies of scale for the program as well as the effect of discontinuing the program.

Some or all of these various issues can serve as the focus of a particular evaluation study. All should be addressed in a comprehensive program evaluation study. It may not, of course, be always desirable or indeed practical to provide in-depth answers to each evaluation question. For example, program evaluation in government carried out under the Treasury Board policy calls for each question to be seriously considered for analysis and investigation, and an appropriate evaluation study to be designed, taking into account the needs of the client (the deputy head), the priority of the evaluation issues, the expected benefits in each case, and the resources available. Conversely, evaluation studies called for by Cabinet or Parliament would likely focus only on the specific issues of interest to Cabinet or Parliament.

Answers to the questions in Table 2 indicate the kind of information to be obtained in an evaluation study. As a result, terms of reference for a particular study should be as specific as possible as to what exactly is required, and not be couched in general terms as in the table. For example, objectives for a service-type program would be specific as to the level of service aimed at - the quantity, quality, timeliness, and fairness of the service.¹¹ Of course, depending on the study design, information may not be collected on all questions: a study may not collect data on how well objectives have been achieved if the main focus of interest is on determining the extent and nature of a particular negative unintended effect. What needs to be determined for a particular evaluation study is the type of information that is required to ensure that, in the eyes of the client, the study addresses the right questions. This is the main purpose of the evaluation assessment study, done before the formal study begins.

The type of evaluation issue being considered may also affect the definition of a program for evaluation purposes. An examination of the rationale of a program would clearly focus on the particular program while an assessment of the impacts on a client group may very well be expanded to include all programs affecting the group.

5.3 The Client of the Evaluation Study

In addition to knowing the questions to be addressed and hence the desired focus of the evaluation study, it is critical that the client of the study be identified and the relevant interests be understood. There may also be numerous users of any study and the clients of the study may wish to take the users' concerns into account.

There are a large number of potential clients (and users) of evaluation studies of government programs:

- Cabinet Committees
- media
- Parliament
- program planners
- public
- program managers
- researchers
- senior government managers.

Given the wide diversity of interests and roles of these different clients, an evaluation study of a particular program done for different clients will usually require a different study design, will answer somewhat different questions and will produce different types of findings.¹²

For example, the 1977 evaluation study of the Unemployment Insurance Commission Benefits (Table 1) was done by the Economic Council of Canada. The Council was interested in determining the large-scale effects of changes in the program. The study was aimed at achieving a better understanding of the program so that future program planning could be improved. An evaluation study of the same program by Parliament, on the other hand, might be aimed instead at determining the effects on individuals or groups of recipients of UIC benefits, or at determining the impact of the program on small communities. Another client might be interested in the impact of the program as an income supplement program rather than as an insurance program, or in whether

the same intended effects could be achieved better through the use of a negative income tax scheme. A program manager, on the other hand, might be interested in small-scale effects that could be improved through incremental changes in the program rather than through major program changes.

Program evaluation studies conducted pursuant to the Treasury Board Policy Circular 1977-47 (see Section 4.2) have as their client the deputy head of the appropriate department or agency and while, as mentioned above, such evaluation would entail addressing all questions in Table 2, the particular evaluation studies done would reflect the specific interests and concerns of the deputy head. This would be evident in the form of the questions addressed and, quite probably, in the groups of activities focused on for each study. Parliament, the Cabinet or central agencies, as a client for evaluation studies, may very well be interested in different issues and in examining different groupings of activities. It could be expected, for example, that evaluation for Parliament, the Cabinet or central agencies would involve proportionally more interdepartmental studies than would program evaluation done for individual deputy heads.

5.4 The Evaluation Method

The fourth aspect to be considered in the development of an appropriate evaluation strategy is the evaluation method or technique to be used. The diversity both of types of evaluations and of clients for evaluation studies, requires the use of different evaluation methods, from quantitative analyses to qualitative investigations, from rigorous to less rigorous assessments, as is evident from Table 1.¹³ An evaluation study which seeks to assess the impact of a particular government program may have to use sophisticated analytical techniques. For example, for planners, computer simulation techniques may be required to analyse the impact of regional economic incentives on the Atlantic provinces. On the other hand, an evaluation done for management in departments¹⁴

... may be quite flexible as to methodology. Its objectives are pertinence, accuracy, and credibility with reference to one or more program management issues. Its audience is principally made up of program managers who are not highly sensitive to the social science researcher's demand for experimental rigor, or to the auditor's demand for standards of evidence.

It should be recognized that for certain programs it may be impossible to obtain definitive quantitative answers to some evaluation questions. This does not imply that there is no feasible evaluation method. Less rigorous quantitative approaches or qualitative approaches are almost always available to provide useful insights into the program and its effects and to give at least partial answers to the more difficult evaluation questions.

5.5 The Resources and Time Available

In designing an appropriate evaluation study, it must be recognized that there are limited resources and time available to carry out such a study. These constraints can often best be addressed during an evaluation assessment phase.

Usually, the evaluation study can be designed to meet specified resource and time constraints. This may require a trade-off between the quantity and the quality of the information to be obtained. For example, the findings of an evaluation to determine the effectiveness of a manpower training program would be much less ambiguous if a special-purpose control group were created. This would be more expensive and take much longer than the use of a comparison group derived from existing administrative data. The 1977 evaluation study of the Canada Manpower Training Program (see Table 1) was faced with this trade-off and chose the comparison-group approach, which gave useful but less definitive findings.

Evaluation studies will be more or less costly depending on the depth of inquiry into specific issues, the number of issues addressed and the availability of information. For much evaluation work - especially when the impacts of the program are being investigated - information from outside the program is required. Obtaining this can be costly and time-consuming but may be essential to a useful study.¹⁵

Evaluation studies done for different clients usually entail different resource and time requirements and must incorporate appropriate time constraints. Evaluations done for management often have a short-term perspective and must be finished in time for decisions on program management. Otherwise, no matter how technically good the study, it may be too late to be usable and hence inappropriate.

Similarly, evaluations done for Parliament or central agencies must take into account the time required for the study to be carried out. Evaluations may be asked for when a particular issue is in the public's mind and must be completed in a time acceptable to the client.¹⁶

Evaluations done for researchers and planners often require more time because of their concentration on the impacts of programs. They also tend to make use of more complex analytical techniques and of more specialized expertise.¹⁷

Finally, evaluation studies should be cost-justified. While there is no general rule to determine if an evaluation study will be worth its cost, some attempt can usually be made to see whether the likely benefits from the study are anything near the cost. An evaluation study is not justifiable when possible benefits are judged clearly to be less than the costs that would be incurred.

5.6 The Need to Ensure Credibility

Probably the most important requirement of an evaluation study is that it be credible. Impartiality, accuracy, and understanding must be maintained in an evaluation.¹⁸ The degree to which these are achieved will differ for different studies. Each evaluation study must be credible in the eyes of the client in order for the findings to be seriously considered.

In the context of evaluations done for deputy heads pursuant to the Treasury Board policy on evaluation, credibility of the evaluation will be achieved through a number of mechanisms. Deputy heads determine what is evaluated and how, and are an integral part of the evaluation process. The evaluation units in departments are organizationally independent from program management, giving them the freedom to be more objective. The Office of the Comptroller General plays a role in assuring objectivity and accuracy of evaluation studies by reviewing selected studies. Finally, public review of any evaluation reports, especially if Access of Information legislation is passed, will help to increase the credibility of the reports in the eyes of the public.

For program managers to be satisfied with evaluations of their programs, they have to be familiar with the strengths and weaknesses of the evaluation techniques used. In addition it has to be made evident to them that the evaluators have dealt with the practicalities of their programs in the study.¹⁹

5.7 Summary

The examples in Table 1 and elsewhere in the text illustrate the wide range of possible kinds of issues that can be addressed under the label "program evaluation". The need for appropriate evaluation studies is evident. In particular, the client of the study and the

evaluation questions addressed determine the evaluation strategy that should be adopted and hence the focus, emphasis and ultimately the findings of the study. Designing appropriate evaluation studies requires careful planning before the actual study commences: an evaluation assessment is almost always required. No single evaluation strategy is always better than another. The use of highly sophisticated evaluation techniques does not guarantee a more useful evaluation. Indeed, the reverse has sometimes been the case. What is needed is an appropriate study: one designed for the particular purpose in mind. Only in this way can non-productive and often costly program evaluation efforts be reduced to a minimum.

As a result, the terms of reference for each evaluation study should be carefully developed and include both the specific questions that are to be addressed and an outline of the methods to be used. This will be of crucial importance, to ensure that the clients of the study - deputy heads, Cabinet or Parliament - receive evaluations which they find useful and suited to their needs.²⁰

6. LIMITATIONS AND REALISTIC EXPECTATIONS FOR EVALUATION STUDIES

The many benefits that result from more and better evaluations of government programs were described earlier in Section 3. However, as indicated by Canadian and foreign experience, there are important limitations to the conduct and use of evaluation studies. This section outlines some of the main lessons learned in this evolving area.

Many governments carry out evaluations of their programs. The status and scope of program evaluation in any country is determined by that country's constitutional arrangements and the way in which it has managed government operations in its particular political setting. Many countries and institutions within each country have tried various approaches to evaluation with varying degrees of success. All have found evaluation of effectiveness difficult to implement in a useful

way. For example, in Sweden, a lot of evaluation activity takes place in its agencies and through Royal Commissions. At the centre, however, the National Audit Bureau now concentrates its efforts on systems audits for economy and efficiency rather than on effectiveness evaluations, in an attempt to improve management practices. The British Treasury in its Program Analysis Reviews conducts evaluations which are used occasionally for planning purposes and have little apparent impact on improved management. In the U.S., the General Accounting Office's stated goal is to conduct what it calls "program results reviews" of all programs. In practice, it finds that program rationales are often not sufficiently developed for such review. Nevertheless, much program evaluation is carried on in the U.S. in the various government departments and agencies at both federal and state level.²¹

Certain points emerge from the Canadian and foreign program evaluation experience to date. Expectations for program evaluation are tempered by the following problems:

- . Program Objectives: the difficulty of specifying measurable objectives for all programs and creating valid measures of objective achievement.
- . Methodology: the lack of standard, successful methods for the study of social and economic problems.
- . Resource Constraints: limitations on all categories of resources for the conduct of evaluations.
- . Using Evaluation Findings: difficulties connected with the adoption of and taking action on findings by both managers and policy makers.
- . A Program's Worth: the lack of a final bottom line means that the final judgement is subjective.

6.1 Program Objectives

Effectiveness evaluation activity in the past has tended to concentrate on the analysis of the extent to which program goals and objectives are achieved²² (Table 2). Success with this type of study

has been limited. The evaluator is often unable to do his or her job because of the ill-defined nature of many program objectives, and the program manager often resents attempts to quantify and measure what he or she sees as inherently qualitative and unmeasurable.

The objectives of some public sector programs are simply not measurable (and may never be). In Sweden, the National Audit Bureau has explicitly shifted from an exclusive goals approach, to an assessment of processes and procedures, working co-operatively with managers. The Auditor General of Sweden reflected recently on 10 years of experience in Sweden:²³

When effectiveness auditing began in Sweden great weight was attached to the matter of goals. The auditors were to utilize the goals set up by the politicians as a point of departure for an assessment of the degree of goal fulfilment in the activities of government agencies. Experience has taught us that often this shot simply isn't on the board. One activity will often have several goals to aim at and the goals may be hazy and diffuse. This situation is not likely to change. Often, politicians are unwilling to tie their hands by stating clear and unequivocal goals for their policy. As a rule, therefore, we in Sweden have had to abstain from any assessment of goal fulfilment.... Indeed, we are working on the assumption that in a longer time perspective audit of operations will come to predominate effectiveness auditing in Sweden.

The underlying problem is not that some programs lack objectives. At the policy level every program has objectives although they may be expressed rather generally. However, specific measurable objectives that enable the evaluator to link program objectives and policy objectives in a logical manner are often difficult to define.²⁴ The extreme difficulty of objectives-oriented evaluation was recognized by the Report of the Independent Review Committee on the Office of the Auditor General of Canada (the Wilson Committee) which reported in 1975:²⁵

Present problems in measuring the effectiveness of spending programs revolve around two major questions. First, how does one define in quantifiable terms what Parliament expects to achieve

in enacting new legislation or by approving a new program through voting funds in an appropriation act? Many programs defy quantification. For example, how can one translate the purposes of the Department of External Affairs into quantitative terms that permit objective measurement of performance? While other programs are susceptible to some form of measurement, they may be intended to achieve a number of different objectives. Some are measurable (for example, the reduction of unemployment or improvement in the standard of living) and some are intangible (such as overcoming a feeling of injustice). Even if some of these identification problems can be resolved, there remains a second major question: How can the results achieved be measured? In some cases, present statistical information may provide a reasonable answer; but in others, new statistical techniques will have to be developed to provide a dependable measure of results.

Even where measurability is not a problem, it can be very difficult to develop agreement on what measures correctly represent the stated objectives:²⁶

Indicators may emerge from the political-bureaucratic process which minimize problems or make them more amenable to solutions and which suit the emotional commitments and interests of the decision-makers and those who influence them. Indicators may be selected to promote the adoption of desired policies, to make these policies appear as successful as possible, and to direct public discussion in the desired directions.

Again, assuming that measurability is not a problem, acceptance of programs may depend on different groups perceiving the objectives of the program to be different:²⁷

Legislation requires ambiguity in the statement of goals so that coalitions can be formed in support of it, and each group can believe that the legislation serves its own special purposes. Insisting that the government's purposes be clear may reduce the scope of legislation and thus limit government involvement in social programs.

Also, an emphasis on measurement will turn attention to those things which can readily be measured. The consequent flow of attention and resources to the area may be at the expense of greater

real need elsewhere. What is measurable may not be of most interest. Furthermore, over emphasis on measurable indicators may turn the indicators into the program's objectives.²⁸

There are a number of other reasons why a program's objectives may be deliberately left vague or unstated, as opposed to being impossible or difficult to define in principle. Objectives may be left unspecified for reasons relating, for example, to security or unity within the nation. In other cases the program may be aimed at providing some service for which there is very little known about how to produce the service effectively, even while the program is earnestly desired. Examples are programs designed to promote racial or religious tolerance or to limit some undesirable behaviour. Objectives are left vague to permit flexibility in the implementation and management of the program, and to provide for responses to new demands or changes in the program environment. Resources can be moved to activities which appear to work as more is learned. Finally, governments of all modern democratic countries pay for some programs, usually for moral or ethical reasons, which might appear to many taxpayers to be ineffective. Treatment for some types of mental disorders is a case: any treatment whatever is thought to be better than no treatment. Withholding treatment indicates hopelessness to the patient and seems callous on the part of the professionals involved. Highlighting the fact that treatment cannot be reasonably demonstrated to be effective may not be a useful contribution to the problem. But the general social objectives may be to indicate concern for such patients and to continue research on this problem, and therefore objective statements will be vague.

Having indicated that for some programs (usually social) a focus on objectives achievement may not be appropriate, two points must be made clear. First, for many other programs, efforts devoted to clarifying and measuring progress towards program objectives will be a very useful evaluation focus. Not doing objectives achievement analysis should be the exception, not the rule. Secondly, even if

focusing on objectives is not appropriate in a particular case, this in no way should limit addressing other evaluation issues. In particular, in such cases, it becomes essential to determine just what are the results of the program.

6.2 Methodology

A common problem and an important limitation to the evaluation of a program's effectiveness is the difficulty of isolating the effects or impact of the program from those of other programs or other factors.²⁹ At any one time our society is subject to a large number of influences. For technical or ethical reasons, few of these can be adequately controlled in order that the effects of the program under study can be unambiguously determined. The difficulty can apply to programs which appear to be quite limited in scope. In 1955, Connecticut instituted a crackdown on speeding. During the year there were 284 traffic deaths compared to 324 in the previous year. Was the increased enforcement responsible for this effect? It may have been in part, or in whole; but, because of many other influences such as weather conditions, the use of seat belts or other safety features it would be most difficult to determine the exact contribution of each factor to the reduced death rate.³⁰

The problem of uncontrolled outside factors manifests itself in other ways. Some government programs involve making grants or contributions to other organizations who then carry out their own project or program alone. Some organizations may refuse to accept funds if one of the conditions for doing so is the provision of information on their clients for program evaluation purposes. Such a dilemma requires a subjective judgement on whether to make the grant or contribution. If the judgement is that a grant ought to be made in any case, then the success or failure of the grant or contribution program will not be able to be evaluated by the government using information on the client population.

When the program being evaluated is composed of smaller components or projects it may not make sense to aggregate the findings of separate reports. Furthermore, at the level of the larger program, the kind of effectiveness questions asked may be quite different.³¹ For example, for a project-based program the effectiveness of each project can be investigated: what are the effects produced and to what extent are objectives achieved? However, the effectiveness of the program as a whole may relate to determining what would have happened without the program. How many projects would have gone on in any event? This is a different kind of evaluation than that of summing up the findings of evaluations of the individual projects. The success of aggregating a number of smaller evaluation studies will vary with each situation.

The dynamic nature of many programs further frustrates the attempt to conduct in-depth evaluations of programs. The single program or the program environment may not remain stable long enough to allow a definitive evaluation study to be carried out.³²

Information systems which have been developed for administrative purposes may not be suitable for in-depth evaluation analysis. Historical information may not exist. In addition to the often high cost of obtaining information on the program, data may not be available for reasons of protection of the privacy of citizens - no matter how otherwise desirable it may be.³³ For example, the use of personal income tax and personal medical records for the evaluation of many programs may be strongly questioned by citizens concerned with the right to privacy.

6.3 Resource Constraints

The conduct of evaluation studies is often costly and time-consuming. Adequate resources may not be available to conduct a credible study. In such cases either the resources must be increased or the study curtailed or postponed.

Program evaluation is a developing area and is not a well-established body of formal knowledge. This imposes another kind of resource constraint. There is a general lack of relevant program evaluation experience in government. The increased application of program evaluation will require training for both managers and practitioners.

A more complex constraint lies in the multi-disciplinary nature of some evaluation studies. Evaluation activities draw from many fields and disciplines such as accounting, economics, education, industrial engineering, the management sciences, political science, medicine, psychology, sociology, statistics and urban affairs.³⁴ Because each program or problem will demand a somewhat different combination of skills, an evaluator may be out of his/her depth in some aspect of the study and the quality of the evaluation could suffer.³⁵ For example, most economists or auditors charged with evaluating the New Horizons program for people retired from the labour force could miss or misinterpret information concerning changes in attitudes and behaviour. Such information might be more properly collected and interpreted by other types of social scientists.

6.4 Using Evaluation Findings

The uses of evaluation findings were outlined earlier in Section 3. A particular evaluation study may, of course, be either more useful or less useful to its client. Although different clients will be using evaluation results with different purposes in mind, for all clients and potential users of evaluation studies there are limitations to using the findings.

Part of the problem is that evaluation studies conducted for one client usually are not well suited for use by others; they do not address the relevant questions. For example, evaluations designed for

policy planners will usually deal in comparative analysis, merely touching on individual program areas of importance to managers, and hence may be of little use for program managers.³⁶

Often, evaluation studies are used by persons other than the original client.³⁷ This is inevitable, but the study should make it clear what issues are and are not addressed. Of course, if a study has not addressed the issues of relevance to the client it is unlikely to be of much use.

Users should also realize that evaluation studies may not produce uncontested findings and may not unequivocally lead to clear recommendations. Particularly in the area of social programs, practitioners acknowledge the difficulty of obtaining results from effectiveness evaluation studies which clearly recommend one feasible course of action. Summary findings tend to be ambiguous to the layman and controversial to the expert. Applied social science research is sometimes not compatible with political or management needs for information on government programs.

In addition, since evaluation studies focus on the impacts of programs, certain findings may have political implications which have not been addressed in the study. In politically sensitive areas this could limit the effective implementation of some recommendations. Related is the fact that the client of the study may not have complete control over taking action on the findings and recommendations. To change a government program often involves several actors who may not agree on the interpretation of the findings or indeed on the focus of the study: from their viewpoint it may have missed key issues. Deputy heads, for example, might be unable to make effective use of the results of an evaluation study if program changes were to have interdepartmental impacts.

Evaluation findings and conclusions will sometimes not be acted upon as recommended for different reasons. There are fairly serious limits on how much, and how rapid, change a program can and should tolerate. Managers must be able to count on some financial commitment to their programs, especially those which represent significant interventions in human lives. Whether or not the program results are those that were intended by the program, they may be very important in lives of individuals. It can also be difficult to obtain competent staff to continue to work in programs where shifts are rapid and unexpected; people develop loyalties to programs and those who have a choice may take their talents out of the Public Service. Consequently, politicians and managers alike may often prefer to shift activities and work for broad continuity and incremental innovation, even though an evaluation report suggests that a more radical change would be more effective.

The usefulness of evaluation studies can be greatly increased if the findings are written and communicated in straightforward language. Too often evaluation studies are written in over-specialized language understandable only to the evaluators and a limited professional community. Clear presentation can help decision-makers better understand the study and its findings.³⁸

6.5 Determining a Program's Worth

The term "program evaluation" could imply that the evaluation study will determine or comment upon the real value of the program - its ultimate worth. This will rarely be the case. Program evaluation is simply an aid to decision-makers. Program evaluation cannot and should not replace the final judgement of the responsible decision-maker. H.S. Havens, the Director of the Program Analysis Division of the General Accounting Office in the United States, has stated this point succinctly.³⁹

Evaluations are not black boxes into which funds can be poured and optimum decisions come out. Only the political process can judge the value of program results and compare their priority to the results of other programs competing for a limited budget.

Most programs have multiple objectives, and there are unavoidable limits to how far the information can be aggregated into summaries to explain, in a single bottom line, whether things are going well or badly. Douglas Hartle, a former Deputy Secretary of the Treasury Board, suggests that this would occur in all cases where programs have multiple objectives. For example, if the objectives of a government were to increase GNP, reduce unemployment and reduce inflation, it would be a simple matter to measure progress toward these individual objectives. It would, however, be very difficult to determine any general improvement in the total economic condition if one of these indicators showed improvement and the other two indicated deterioration.⁴⁰ This is the job of the appropriate decision-makers, although the evaluator may be asked to express his or her opinion and to make recommendations.

Even were a program to have a single objective which it had achieved, it does not necessarily follow that the program has been worth while. The objective may have been inappropriate, the program may have created a number of negative unintended impacts and effects or it may have been possible to achieve the objective at lower cost. The evaluation study should consider all these possibilities, and provide the decision-makers with the relevant information so that they can judge the worth of the program.

7. WHAT MAKES AN EVALUATION STUDY SUCCESSFUL?

Experience gained in program evaluation in Canada and elsewhere should be recognized and taken into account in developing new initiatives. This experience suggests that a number of aspects of successful evaluations can be identified:

- . The program being evaluated should be carefully identified and described.
- . An evaluation assessment - a pre-evaluation planning phase - should be carried out to determine the appropriate focus and methodology for the evaluation study.
- . The evaluation study should have specific terms of reference which include, among other things, the specific evaluation questions to be addressed, so that the particular objective of the study is well understood.
- . It should be expected that the potential benefits of carrying out an evaluation would outweigh the cost.
- . Evaluation studies should involve the manager of the program being evaluated and, to enhance objectivity, should involve independent evaluators.
- . The evaluation study should be timely in the sense of being completed in time for its consideration in the appropriate decision-making context.
- . There should be a sufficient mix of resources of the right quality available for the particular evaluation study.
- . The methodology used should be credible to the client.
- . Expectations as to the possible findings of the evaluation study should be realistic.
- . The language of evaluation reports should not be over-specialized.
- . The evaluation report should clearly describe what the specific questions to be addressed were, how they were analyzed and investigated, what evidence was collected and what conclusions followed from this evidence.
- . The evaluation study should have a specific client who has a commitment to examining the findings.

FOOTNOTES

1. Office of the Comptroller General, Internal Audit and Program Evaluation in the Government of Canada, August 7, 1979. pp 2-3.
2. The Glassco Commission stated: "Just as the changing scope and character of federal activity must respond to the changing character and needs of the Canadian community, so the government's structure and methods must take account of new resources and techniques ... budgetary and accounting systems which permit better control over the allocation of financial resources and the assessment of performance; operational research and other techniques for evaluating alternative courses of action and for designing and appraising methods and systems." The Royal Commission on Government Organization; Management of the Public Service, Vol. 1. (Ottawa: The Queen's Printer, 1962), pp. 45-6.
3. Government of Canada, Estimates for the Fiscal Year Ending March 31, 1978, Introduction, p. 1-6.
4. "One respect in which a PPB System is greatly different from traditional forms of government budgeting is in its concentration on the results or output and benefits as opposed to just a consideration of the resources required. Intensive study is made of feasible alternative ways of attaining defined objectives with a view to determining the approach which is most likely to achieve the greatest benefit for a given cost or conversely, the approach by which a given objective will be achieved at minimum cost". Government of Canada, Planning Programming Budgeting Guide. (Ottawa: The Queen's Printer, 1969), p. 4.
5. In Office of the Comptroller General, op. cit., p. 2, internal audit is distinguished from program evaluation and is defined as "...the systematic, independent review and appraisal of all departmental operations, including administrative activities, for purposes of advising management as to the efficiency, economy and effectiveness of the internal management practices and controls".
6. Aaron Wildavsky and Jeanne Nienaber, The Budgeting and Evaluation of Federal Recreation Programs. (New York: Basic Books, 1973), p. xvi.
7. Formal planning for evaluation studies has become an accepted part of the evaluation process. See, for example, Michael Q. Patton, Utilization Focused Evaluation. (Beverly Hills, California: Sage, 1978) and Leonard Rutman, Planning Useful Evaluations, Evaluability Assessment. (Beverly Hills, California: Sage, 1980) for two approaches (in the U.S. context) to such planning. The approach discussed in this paper - evaluation assessment - is described more fully in the two OCG documents referred at the beginning of this discussion paper.

8. Disentangling the various levels of a program is a critical stage in determining the purpose and type of evaluation. See United States General Accounting Office, Comprehensive Approach for Planning and Conducting a Program Results Review. (Washington, D.C.: General Accounting Office, 1978).
9. For an explanation of why new programs need time to mature before they are submitted to the scrutiny of evaluation, see: Daniel P. Moynihan, "Policy vs. program in the 70's", Public Interest, No. 20 (Summer, 1970) pp. 90-100.
10. For example, Malcolm Parlett and David Hamilton envisage a smaller role for comparative experimental design in the evaluation of programs that are unique and innovative. See their, "Evaluation as illumination: a new approach to the study of innovatory programs," in Gene V. Glass (ed.), Evaluation Studies Review Annual. (Beverly Hills, California: Sage Publications, 1976), pp. 140-57.
11. Level-of-service issues are a legitimate part of a program evaluation. An internal audit concentrates on the systems and procedures for providing the service, while an evaluation study is concerned with results: in this case the service levels produced as seen by the clients.
12. Differing perspectives for evaluation studies are discussed in E. Chelimsky (ed.), Proceedings of a Symposium on the Use of Evaluation by Federal Agencies. Vol. II. An Analysis of the Symposium Proceedings. (MacLean, Va: MITRE Corp., 1977).
13. For a discussion of evaluation's differing techniques and strategies see the "Introduction" of Gene V. Glass (ed.), Evaluation Studies Review Annual. (Beverly Hills, California: Sage Publications, 1976). pp. 9-14 and L.S. Kogan and A.W. Shyne, "Tender-minded and tough-minded approaches in evaluation research," International Social Science Bulletin, Vol. 4, No. 2, 1966, pp. 12-17.
14. Thomas Kelly, Program Evaluation: The Management Perspective. Paper presented at the ORSA/TIMS Joint Meeting, Atlanta, Georgia, November 8, 1977, p. 8.
15. Evaluation creates pressures to use available data. The pitfalls of information collection for evaluation purposes are discussed by Louis A. Ferman, "Some perspectives on evaluation of social welfare programs," The Annals of the American Academy of Political and Social Sciences, Vol. 385, 1969, pp. 143-156.
16. How an issue will rise to public view, capture centre stage for a time and then fade away is outlined by Anthony Downs, The Issue-Attention Cycle and Improving Our Environment. (Chicago: Real Estate Research Corporation, 1971).

17. See Peter H. Rossi, "Booby traps and pitfalls in the evaluation of social action programs," in F.G. Caro (ed.), Readings in Evaluation Research, Second Edition. (Beverly Hills: Sage Publications, 1977), pp. 239-252.
18. Balancing bias is the scope of an article by Michael Scriven, "Evaluation bias and its control," in Gene V. Glass (ed.), Evaluation Studies Review Annual. (Beverly Hills, California: Sage Publications, 1976), pp. 119-139. Also see: Marcia Guttentag, "Subjectivity and its use in evaluation research" Evaluation, Vol. 1, No. 2, 1973, pp. 60-65.
19. For a program management perspective, see: Joseph S. Wholey, "What can we actually get from program evaluation?", Policy Sciences Vol. 3, 1972, pp. 361-369.
20. The importance of the design of an evaluation for legislative consumption is discussed by Allen Schick, "Congress and the details of administration," Public Administration Review, September, 1976, pp. 516-527. See also Herbert Roback, "Program evaluation by and for Congress", The Bureaucrats, Vol. 5, No. 1, 1976, pp. 11-36.
21. For a discussion of the respective countries see (a) Sven I. Ivarsson, Effectiveness Control In A Program Budgeting System. (Stockholm: National Audit Bureau, 1974), (b) I. Galnoor, "Reforms of public expenditures in Great Britain," Canadian Public Administration, Vol. 17, No. 2, (Summer, 1974) pp. 289-320 and (c) the Controller General of the United States, Evaluation and Analysis to Support Decision-Making. (Washington, D.C.: General Accounting Office, 1976). For a discussion comparing evaluation efforts in several countries see H. Jamous, "Evaluation of Government Programmes: Reflections for a Debate", prepared for the OECD Committee for Scientific and Technological Policy, SPT(80)3, February 1980 and Robert A. Levine, "An Overview of Program Evaluation and Policy Analysis in Western Nations", Report on a Pilot Project Sponsored by the German Marshall Fund of the United States, April, 1980.
22. In this discussion paper, no distinction is made between goals and objectives.
23. G. Rune Berggren, "Effectiveness Auditing in Swedish National Administration," International Journal of Government Auditing, April 1978, pp. 5-6.
24. This point is discussed by W. Irwin Gillespie, "Fools Gold: The Quest for a Method of Evaluating Government Spending", in C. Bruce Doern and Allan M. Maslove(eds.), The Public Evaluation of Government Spending. (Montreal: Institute for Research on Public Policy, 1979).

25. Office of the Auditor General of Canada, Report of the Independent Review Committee (Ottawa: Information Canada, 1975), p. 34.
26. Allan M. Maslove, "Indicators and policy formation, "Canadian Public Administration, Vol. 18, No. 3, (Fall, 1975) p. 483. See also M.J. Moffat and T.E. Reid, "Comment", Canadian Public Administration Vol. 19, No. 4, (Winter, 1976) and A. Maslove, "A Reply" Canadian Public Administration Vol. 20, No. 3, (Fall, 1977).
27. M. Rein and S.H. White, "Can policy research help policy?" The Public Interest, (Fall, 1977) p. 123.
28. This point is discussed, for example, by Irwin Deutscher, "Toward Avoiding the Goal-Trap in Evaluation Research", in The Evaluation of Social Programs, C.C. Abt. (ed.) (Beverly Hills: Sage Publications, 1976). pp. 249-268.
29. Isolating the impacts of government programs is discussed by James S. Coleman, "Problems of conceptualization and measurement in studying policy impacts," in K.M. Dolbeare (ed.), Public Policy Evaluation. (Beverly Hills: Sage Publications, 1975), pp. 19-40.
30. Donald T. Campbell, "Reforms as experiments," in F.G. Caro, ed., Readings in Evaluation Research. (New York: Sage, 1977), p. 177.
31. Goal fulfilment occurs at different levels in a program. Aggregating lower-level goal achievements in a cumulative way may not represent the over-all effectiveness of the program. See: S.I. Ivarsson, op cit, pp. 7-9.
32. V.G. Birnberg and N.M. Gandhi explain that social programs go through various stages of development and suggest that evaluation should match the appropriate stage. See their "The accountants are coming: how accountants can help policy makers in social program evaluation", Policy Science, Vol. 8, 1977, pp. 468-481.
33. Ferman, op. cit.
34. Robert Perloff, Evelyn Perloff, and Edward Sussna, "Program evaluation," Annual Review of Psychology (1976) p. 571.
35. Gerald Gordon and Edward V. Morse, "Evaluation research," Annual Review of Sociology, Vol. 1, 1975 found in their examination of three years of reports on evaluation activity in the Sociological Abstracts that only 7 of 93 studies reviewed had a good grasp of statistical foundations.
36. On a program manager's information needs see Victor G. Nielsen, "Why evaluation does not improve program effectiveness," Policy Studies Journal, Vol. 3, No. 4, (Summer, 1975).

37. The implications of various actors recognizing the uses of evaluations is discussed by R.E. Floden and S.S. Weiner, "Rationality to ritual: the multiple roles of evaluation in governmental processes", Policy Sciences Vol. 9, 1978, pp. 9-18.
38. The language of evaluation and other problems associated with implementing evaluation findings are examined by H.R. Davis and S.E. Salasin, "The utilization of evaluation", in Marcia Guttentag and Elmer Struening (eds.). Handbook of Evaluation Research, Volume I. (Beverly Hills: Sage Publications, 1975) pp. 621-666.
39. H.S. Havens, Excerpt from a statement made before the Senate Committee on Human Resources, Washington, D.C., October 6, 1977.
40. Douglas G. Hartle, "A proposed system of program and policy evaluation," Canadian Public Administration, Vol 16, No. 2 (Summer, 1973).

32. V.G. Birnberg et N.M. Gandhi expliquent que les programmes sociaux se déroulent en diverses étapes et ils proposent leur évaluation ait lieu à l'étape appropriée. Consulter leur article "The Accountants Are Coming: How Accountants Can Help Policy Makers In Social Program Evaluation", Policy Science, Vol. 8, 1977 pp. 468 à 481.
33. Ferman op cit.
34. Robert Perloff, Evelyn Perloff et Edward Sussna, "Program Evaluation", Annual Review of Psychology (1976) p. 571.
35. Gerald Gordon et Edward V. Morse, "Evaluation Research", Annual Review of Sociology vol. 1 (1975) ont découvert dans l'étude qu'ils ont menée pendant trois ans sur les rapports concernant l'évaluation et dont ils ont rendu compte dans Sociological Abstracts que, sur les 93 études examinées, seulement 7 reposaient sur une base statistique solide.
36. Consulter l'article de Victor G. Nielsen, "Why Evaluation Does Not Improve Program Effectiveness", Policy Studies Journal, Vol. 3, n° 4 (été 1975) p. 386 sur les besoins d'information.
37. R.E. Floden et S.S. Weiner, étudient les répercussions de la reconnaissance, par divers clients, des utilisations des évaluations effectuées dans "Rationality to ritual: the multiple roles of evaluation in governmental processes", Policy Sciences Vol 9, (1978), pp. 8 à 18.
38. La terminologie relative à l'évaluation et autres problèmes liés à la mise en oeuvre des conclusions découlant de l'évaluation sont examinées par H.R. Davis et S.E. Salasin dans "The utilization of evaluation", in Marcia Gutentag et Elmer Struening (ed.) Handbook of Evaluation Research, vol 1. (Beverly Hills: Sage Publications, 1975) pp. 621 à 666.
39. H.S. Havens, extrait d'une déclaration prononcée devant le comité sénatorial sur les ressources humaines à Washington D.C., le 6 octobre 1977.
40. Douglas G. Hartle, "A proposed system of program and policy evaluation", Administration publique du Canada, Vol. 16, N° 2 (été 1973).

22. Dans le présent document, nous ne faisons pas de distinction entre les buts et les objectifs.
23. G. Rune Berggren, "Effectiveness Auditing in Swedish National Administration, International Journal of Government Auditing (avril 1978) pp. 5 et 6.
24. Ce point est présenté par W. Irwin Gillespie dans "Fools Gold: The Quest for a Method of Evaluating Government Spending", dans The Public Evaluation of Government Spending, C. Bruce Doern et Allan M. Maslove (éd.) (Montréal, Institut de recherches politiques, 1979).
25. Bureau du Vérificateur général du Canada, Rapport du Comité indépendant sur les fonctions du Vérificateur général du Canada, (Ottawa: Information Canada, 1975) p. 34.
26. Allan M. Maslove, "Indicators and policy formation", Administration publique du Canada, vol. 18, n. 3 (automne 1975) p. 483. A consulter aussi: M.J. Moffat et T.E. Reid, "Comment", Administration publique du Canada, vol. 19, n. 4, (hiver, 1976) et A. Maslove "A Reply" Administration publique du Canada, vol. 20, n. 3, (automne, 1977).
27. M. Rein et S.H. White, "Can policy research help policy?" The Public Interest (automne, 1977) p. 123.
28. Cette question est traitée, par exemple, par Irwin Deutscher, "Toward Avoiding the Goal-Trap in Evaluation Research", in The Evaluation of Social Programs, C.C. Abt. (ed.) (Beverly Hills: Sage Publications, 1976). pp. 249-268.
29. James S. Coleman examine les façons d'isoler les répercussions des programmes gouvernementaux dans l'ouvrage suivant: "Problems of conceptualization and measurement in studying policy impacts", (éditions K.M. Dobeare), Public Policy Evaluation, (Beverly Hills: Sage Publications, 1975) pp. 19 à 40.
30. Donald T. Campbell, "Reforms as experiments", (éditions F.G. Caro) Readings in Evaluation Research. (New York: Sage, 1977) p. 177.
31. Dans le cadre d'un programme, les objectifs s'atteignent à différents niveaux. Le regroupement des objectifs de moindre importance peut ne pas refléter l'efficacité globale du programme. Voir S.I. Ivarsson, op. cit. pp. 7 à 9.

14. Thomas Kelly, Program Evaluation: The Management Perspective. Document présenté lors du Joint Meeting de ORSA/TIMS, à Atlanta en Georgie, le 8 novembre 1977, p. 8.
15. L'évaluation contrainte à employer les données disponibles. Louis A. Ferman examine les pièges de la collecte des données qui serviront aux évaluations, "Some perspectives on evaluation of social welfare programs", The Annals of the American Academy of Political and Social Sciences, Vol. 385, 1969, pp. 143-156.
16. Anthony Downs décrit la façon dont une question attire l'attention du public, l'accapare pendant un certain temps, puis disparaît progressivement. The Issue - Attention Cycle and Improving Our Environment. (Chicago: Real Estate Research Corporation, 1971)
17. Lire Peter H. Rossi, "Booby traps and pitfalls in the evaluation of social action programs", qui fait partie du volume de F.G. Caro (éd.), Readings in Evaluation Research, Deuxième édition. (Beverly Hills: Sage Publications, 1977) pp. 239-252.
18. Le rétablissement de la partialité constitue le cadre d'un article de Michael Scriven, "Evaluation Bias and its Control", (éd. Gene V. Glass), Evaluation Studies Review Annual. (Beverly Hills, California: Sage Publications, 1976). pp. 119 à 139. A consulter également: Marcia Gutentag, "Subjectivity and its use in evaluation research" Evaluation, 1973 Vol 1, N° 2, pp. 60 à 65.
19. Du point de vue de la gestion, se référer à: Joseph S. Wholey, "What can we actually get from program evaluation", Policy Sciences Vol 3, (1972) pp. 361 à 369.
20. Allen Schick discute de l'importance que revêt l'élaboration d'une évaluation à l'intention des législateurs, "Congress and the details of administration", Public Administration Review, septembre 1976, pp. 516-527. Lire également Herbert Roback, "Program evaluation by and for Congress", The Bureaucrats, Vol. 5, N° 1, 1976, pp. 11-36.
21. Afin de connaître la situation dans différents pays, consulter les ouvrages suivants a) Sven I. Ivarsson, Effectiveness Control In A Program Budgeting System, (Stockholm: National Audit Bureau, 1974) b) I. Galinor, "Reforms of Public Expenditures in Great Britain", Administration publique du Canada, Vol. 17, N° 2, pp. 289 à 320 et c) Controller General of the United States, Evaluation and Analysis to Support Decision Making. (Washington, D.C.: General Accounting Office, 1976). Pour obtenir une étude comparée des efforts qui ont été faits dans plusieurs pays en ce qui concerne l'évaluation, voir H. Jamous, Evaluation of Government Programs: Reflections for a Debate, préparé par le Comité de l'O.C.D.E. chargé des politiques

7. La planification officielle des évaluations est devenue une partie reconnue du processus de l'évaluation. Voir Michael Q. Patton, Utilization Focused Evaluation (Beverly Hills, California: Sage, 1978) et Leonard Rutman, Planning Useful Evaluations, Evaluability Assessment (Beverly Hills, California: Sage, 1980) au sujet de deux méthodes employées aux Etats-Unis pour cette planification. La méthode examinée dans le présent document (étude préparatoire d'une évaluation) est énoncée d'une manière plus complète dans les deux documents du BCG mentionnés au début du présent exposé.
8. La compréhension des différents niveaux d'un programme constitue un stage critique dans la détermination du but et du genre d'évaluation. Comprehensive Approach for Planning and Conducting a Program Results Review, du United States General Accounting Office. (Washington, D.C.: General Accounting Office, 1978).
9. Pour obtenir une explication des raisons pour lesquelles on doit laisser mûrir les nouveaux programmes avant de les soumettre à une évaluation, voir: Daniel P. Moynihan "Policy vs. program in 70's" Public Interest No. 20 (été, 1970) pp. 90 à 100.
10. Malcolm Parlett et David Hamilton, par exemple, envisagent un rôle moins important pour la conception comparative expérimentale dans l'évaluation des programmes unifiés et innovateurs. Voir leur "Evaluations as illumination: a new approach to the study of innovative program", (édition Gene V. Glass), Evaluation Studies Review Annual. (Beverly Hills, California: State Publications, 1976) pp. 140 à 157.
11. La question des niveaux de service fait partie intégrante de l'évaluation de l'efficacité. La vérification interne porte essentiellement sur les systèmes et méthodes permettant de fournir le service alors que l'étude d'évaluation s'intéresse, comme toujours, aux résultats; dans le cas présent, il s'agit, comme résultat, du niveau des services fournis, du point de vue des clients.
12. Il est question des conceptions différentes des études d'évaluation dans E. Chelimsky (ed.), Proceedings of a Symposium on the Use of Evaluation by Federal Agencies, Vol. II. An Analysis of the Symposium Proceedings. (Maclean, va: MITRE Corp., 1977).
13. Pour un examen des différentes techniques et méthodes d'évaluation, lire l'introduction de Gene V. Glass (ed.), Evaluation Studies Review Annual. (Beverly Hills: California: Publications Sage, 1976) pp. 9-14 et L.S. Kogan et A.W. Shyne, "Tender-minded and tough-minded approaches in evaluation research", International Social Science Bulletin, 1966, Vol. 4, No. 2, pp. 12-17.

NOTES

1. Bureau du Contrôleur général, La vérification interne et l'évaluation des programmes au sein du gouvernement du Canada, 7 août 1979, pp. 2-3
2. Voici ce que la Commission Glassco a dit à ce sujet: "Tout comme la portée et le caractère changeants de l'activité fédérale doivent répondre au caractère et aux besoins changeants de la population canadienne, le mode d'organisation du gouvernement et ses méthodes doivent profiter des ressources et techniques nouvelles... méthodes budgétaires et systèmes de comptabilité permettant de mieux contrôler la répartition des ressources financières et la détermination du rendement, recherche opérationnelle et autres méthodes pour l'étude de solutions de rechange et pour la mise au point et l'évaluation de méthodes et de systèmes. La Commission royale d'enquête sur l'organisation du gouvernement; La Gestion de la Fonction publique, Volume 1. (Ottawa: L'Imprimeur de la Reine, 1962) pp. 47 à 50.
3. Le Gouvernement du Canada, Budget des dépenses pour l'année financière se terminant le 31 mars 1978, Introduction, pp. 1-7.
4. "Le système PPB diffère grandement des formes traditionnelles de budgétisation gouvernementale en ce qu'il porte sur les résultats ou "sorties" et sur les bénéfices plutôt que sur une simple étude des ressources nécessaires. On y procède à une étude poussée des différents moyens possibles d'atteindre des objectifs précis afin de déterminer la méthode qui fera réaliser le plus gros bénéfice par rapport à un coût donné ou, inversement, celle qui permettra d'atteindre un objectif déterminé à un coût minimum. Gouvernement du Canada, Guide de Planification, programmation, budgétisation (Ottawa, Imprimerie de la Reine, 1969), p. 4.
5. Dans le Bureau du Contrôleur général, op. cit. p. 2, la vérification interne est décrite comme étant distincte de l'évaluation des programmes et elle y est définie comme étant l'examen et l'appréciation systématiques et indépendants de toutes les opérations du ministère, y compris les activités administratives, en vue de conseiller la direction sur l'efficacité, l'économie et l'efficacité des pratiques et des contrôles internes de gestion.
6. Aaron Wildavsky et Jeanne Nienaber, The Budgeting and Evaluation of Federal Recreation Programs. New York: Basic Books, 1973 p. xvi.

- Il faut que la méthodologie utilisée impose sa crédibilité au client.
- Il faut envisager de façon réaliste les conclusions possibles de l'évaluation.
- Les rapports d'évaluation ne doivent pas être rédigés dans un style trop spécialisé.
- Le rapport d'évaluation doit décrire clairement quelles étaient les questions particulières soulevées, comment elles ont été analysées et examinées, quelles sont les données qui ont été recueillies et quelles sont les conclusions qui en découlent.
- L'évaluation doit être faite pour un client précis qui s'est engagé à en étudier les résultats.

Même si un programme n'a qu'un seul objectif qu'il a atteint, il n'en découle pas nécessairement que le programme soit valable. L'objectif peut ne pas être pertinent, le programme peut avoir entraîné plusieurs répercussions et effets négatifs non intentionnels ou encore il peut être possible d'atteindre l'objectif à un coût inférieur. L'étude d'évaluation devrait soupeser toutes ces possibilités et fournir aux décisionnaires les renseignements pertinents afin qu'ils puissent établir la valeur du programme.

7. QU'EST-CE QUI CONTRIBUE AU SUCCÈS D'UNE ÉVALUATION?

L'expérience acquise dans le domaine de l'évaluation des programmes au Canada et ailleurs doit être reconnue et prise en considération lors de l'élaboration de nouveaux projets. Cette expérience permet de déterminer certains facteurs qui contribuent à l'efficacité des évaluations:

- Il faut soigneusement déterminer et définir le programme qui fait l'objet d'une évaluation.
- Afin de déterminer le point sur lequel doit se concentrer l'évaluation ainsi que la méthodologie à employer, on fera une étude préparatoire de cette évaluation. Il s'agit d'une étape de planification qui précède l'évaluation proprement dite.
- Il faut préciser le mandat de l'évaluation lequel doit inclure, entre autres, les questions précises à poser, afin que l'objectif particulier de l'étude soit bien compris.
- L'évaluation doit comporter des avantages éventuels qui en justifient le coût.
- Les évaluations devraient impliquer les gestionnaires du programme que l'on évalue et, pour assurer leur objectivité, devraient être faites par des évaluateurs autonomes.
- L'évaluation doit être terminée au moment opportun pour pouvoir être étudiée dans le contexte nécessaire à la prise de décision.
- Il faut disposer de ressources suffisantes et pertinentes pour effectuer l'évaluation.

6.5 L'établissement de la valeur d'un programme

L'expression "évaluation de programme" pourrait laisser croire qu'il s'agit de déterminer ou d'expliquer la valeur réelle d'un programme, c'est-à-dire sa valeur absolue. Pourtant, c'est rarement le cas. L'évaluation de programme est un instrument destiné aux décisionnaires. Elle ne peut, ni doit, remplacer la décision finale prise par ces derniers. H.S. Havens, directeur de la Division de l'analyse des programmes du General Accounting Office des Etats-Unis a illustré ce point succinctement:

Les évaluations ne sont pas des boîtes noires dans lesquelles les fonds entrent d'un côté ressortent en décisions optimales de l'autre. Seul le processus politique permet de déterminer la valeur des résultats d'un programme et de comparer leur importance à celle d'autres programmes, qui cherchent tous à obtenir une part d'un budget limitée.

Les programmes ont le plus souvent plusieurs objectifs. C'est pourquoi, inévitablement, on en arrive à un point où il est impossible de condenser davantage l'information et de résumer la situation en une "seule ligne", à savoir si les choses vont bien ou mal. D'après Douglas Hartle, ancien sous-secrétaire du Conseil du Trésor du Canada, cela se produit chaque fois que les programmes visent des objectifs multiples. Par exemple, si un gouvernement se fixait comme objectifs d'augmenter le PNB, de diminuer le chômage et l'inflation, il serait tout à fait simple d'évaluer les progrès réalisés dans la voie de la concrétisation de chaque objectif. Par contre, il serait extrêmement difficile de déterminer si la situation économique globale s'est généralement améliorée lorsque l'un de ces indicateurs marque une amélioration de la situation tandis que les deux autres indiquent une détérioration⁴⁰. C'est là qu'interviennent les décisionnaires appropriés, bien que l'on puisse demander à l'évaluateur d'exprimer son opinion et de faire des recommandations.

À cela, il faut ajouter que le client de l'étude peut ne pas exercer un contrôle absolu sur l'application des conclusions et des recommandations de l'étude. L'évolution des programmes gouvernementaux met souvent en cause plusieurs agents qui peuvent ne pas s'entendre sur l'interprétation des conclusions ni même sur l'objet fondamental de l'étude: de leur point de vue, celle-ci est peut-être passée à côté de questions cruciales. Les sous-chefs, par exemple, peuvent être incapables d'utiliser efficacement les résultats d'une étude d'évaluation si les modifications à apporter aux programmes comportent des incidences interministérielles.

Parfois, il ne sera pas donné suite aux résultats et aux conclusions d'une évaluation tel que recommandé et ce, pour diverses raisons. Il existe des restrictions inévitables quant à l'importance et à la rapidité des changements qu'un programme peut et doit subir. Les administrateurs devraient pouvoir compter sur un certain engagement financier à l'endroit de leurs programmes, surtout pour ceux qui représentent des interventions importantes dans les vies humaines. Même si les résultats d'un programme ne sont pas ceux qui en avaient été escomptés, ceux-ci peuvent avoir une grande importance dans les vies des personnes en cause. De plus, il peut s'avérer difficile de recruter un personnel compétent lorsque les changements sont rapides et inattendus; certains finissent par s'identifier à des programmes et ceux qui ont le choix pourrnt décider de quitter la fonction publique. En conséquence, les politiciens, tout comme les administrateurs, préféreront souvent modifier certaines activités et travailler dans le sens d'une continuité et d'innovations constantes, même si le rapport d'évaluation laisse supposer qu'un changement plus radical pourrait être plus efficace.

Les évaluations pourrnt être beaucoup plus utiles si leurs résultats étaient communiqués et rédigés dans une langue simple. Trop souvent, elles emploient des termes trop spécialisés, accessibles uniquement aux évaluateurs et à un groupe professionnel restreint. Une présentation claire peut permettre aux cadres responsables de mieux comprendre l'étude et ses résultats.

Une partie du problème réside dans le fait que les études

d'évaluation réalisées pour un client donné ne se prêtent habituellement pas bien à leur utilisation par d'autres puisque les questions pertinentes pour leurs besoins ne sont pas toujours soulevées. Par exemple, les évaluations conçues pour les responsables de la planification des politiques se fondent habituellement sur des analyses comparatives et ne font qu'effleurer les aspects de programmes particuliers qui intéressent les gestionnaires, ce qui fait qu'elles sont de peu d'utilité pour les gestionnaires de programmes ³⁶.

Il arrive souvent que les études d'évaluation sont utilisées par d'autres que le client de l'étude ³⁷. C'est là une chose inévitable, mais l'étude devrait établir clairement quelles sont les questions qui ont été étudiées et celles qui ne l'ont pas été. Bien sûr, si une étude a laissé de côté les questions qui intéressent le client, il est probable qu'elle sera de peu d'utilité.

Les usagers doivent aussi se rendre compte que les études d'évaluation peuvent aboutir à des conclusions contestables et à des recommandations équivoques. Notamment dans le domaine des programmes sociaux, les praticiens reconnaissent la difficulté d'obtenir, à partir des études d'évaluation de l'efficacité, des résultats qui recommandent clairement une ligne d'action pratique. Les résumés des conclusions ont tendance à se révéler ambigus aux yeux des profanes et à susciter des controverses parmi les experts. Il arrive que la recherche appliquée en sciences sociales soit incompatible avec les besoins politiques ou gestionnels de renseignements sur les programmes gouvernementaux.

En outre, puisque les études d'évaluation mettent l'accent sur les répercussions des programmes, certaines conclusions peuvent avoir des incidences politiques qui n'ont pas été soulevées dans l'étude. Dans les domaines qui sont délicats sur le plan politique, cela peut de fait limiter la mise en oeuvre de certaines recommandations.

L'évaluation des programmes est un domaine en pleine évolution et n'est pas une science reconnue; ce fait impose de nouvelles contraintes. L'administration gouvernementale manque généralement d'expérience pertinente en matière d'évaluation de programme. Le recours grandissant à l'évaluation des programmes exigera une certaine formation tant de la part des gestionnaires que des techniciens.

La contrainte que représente la multidisciplinarité de certaines études d'évaluation est encore plus grave. En effet, l'évaluation emprunte ses techniques de nombreuses disciplines dont la comptabilité, l'économie, l'éducation, le génie industriel, les sciences de la gestion, les sciences politiques, la psychologie, la médecine, la sociologie, la statistique et l'urbanisme³⁴. Etant donné que chaque programme ou problème nécessite une combinaison quelque peu différente de ces connaissances, il peut arriver qu'un évaluateur n'ait pas les compétences voulues pour traiter l'un des aspects de l'étude, ce qui risque de nuire à la qualité de l'évaluation³⁵. Par exemple, la plupart des économistes et des vérificateurs financiers chargés d'évaluer le programme Horizons nouveaux, destiné aux personnes ayant quitté les rangs de la population active, pourraient être incapables de repérer les indices d'un changement d'attitude ou de comportement ou de les interpréter correctement. La collecte et l'interprétation de ce type de renseignements devraient, au contraire, être confiées à d'autres spécialistes des sciences sociales.

6.4 L'utilisation des résultats des évaluations

Les utilisations qui sont faites des conclusions des évaluations ont été décrites plus haut dans l'article 3. Une étude d'évaluation particulière peut évidemment se révéler plus ou moins utile pour le client. Bien sûr, les divers clients utiliseront les résultats des évaluations à des fins diverses, mais à l'égard de tous les clients et de tous les utilisateurs éventuels des études d'évaluation, il y a des limites qui s'imposent dans l'utilisation des conclusions des études.

axé sur la mise en oeuvre de projets on peut se demander quels effets ils ont produits et dans quelle mesure les objectifs ont été atteints. Or, dans le cas du programme dans son ensemble, il pourrait s'agir au contraire de déterminer ce qui se serait produit en son absence et combien de projets auraient été mis en marche, en tout état de cause. Cette façon de procéder diffère de celle qui consiste à additionner les résultats des évaluations de différents projets. Bref, la mise en commun d'évaluations de faible envergure est plus ou moins satisfaisante selon la situation.

Le caractère évolutif de nombreux programmes constitue un autre obstacle à leur évaluation scientifique. Il arrive en effet qu'un programme ou le contexte dans lequel il se situe ne soient pas stables assez longtemps pour qu'une évaluation définitive puisse être effectuée³².

Les systèmes d'information mis au point à des fins administratives ne se prêtent pas nécessairement à une évaluation en profondeur. Il se peut qu'il n'existe aucune donnée disponible. De plus, non seulement la collecte de renseignements sur le programme est-elle souvent très coûteuse, mais on ne peut obtenir certaines données, quelques-unes qu'elles puissent être, afin d'assurer la protection de la vie privée des citoyens³³. Par exemple, la consultation des dossiers de l'impôt sur le revenu des particuliers ou de dossiers médicaux personnels pourrait faire l'objet de vives contestations de la part des citoyens soucieux de faire respecter leur droit à la vie privée.

6.3 Les restrictions sur le plan des ressources

Effectuer des études d'évaluation nécessite souvent beaucoup de temps et d'argent. Les ressources suffisantes peuvent ne pas être disponibles pour mener une étude fiable. Dans ces cas, il faut, soit augmenter les ressources, soit abréger ou atermoyer l'étude.

raisons d'ordre technique ou moral, de sorte qu'il est très difficile de déterminer sans ambiguïté l'incidence d'un programme donné. Ce genre de difficulté peut survenir dans le cas de programmes dont la portée paraît plutôt limitée. Par exemple, en 1955, l'Etat du Connecticut a institué un contrôle rigoureux des limites de vitesse. Cette année-là, l'Etat n'a enregistré que 284 pertes de vie sur les routes, contre 324, l'année précédente. L'application plus stricte de la loi était-elle à l'origine de cette baisse? Elle peut l'avoir été totalement ou en partie, mais, en raison de maintes autres influences telles les conditions météorologiques, l'utilisation des ceintures de sécurité et ainsi de suite, il serait très difficile de déterminer la contribution précise de chaque facteur à la réduction du taux de mortalité³⁰.

Le problème que posent les facteurs extérieurs incontrôlables se manifeste d'autres manières. Certains programmes gouvernementaux comportent la prestation de subventions ou de contributions à divers organismes, lesquels se chargent de mettre en oeuvre un programme ou un projet de leur cru. Or, certains organismes refuseront les fonds si l'une des conditions qui s'appliquent à l'octroi de ceux-ci est la communication de renseignements sur leurs clients à des fins d'évaluation du programme. Ce genre de dilemme exige un jugement subjectif quant à la question de décider si l'on accordera ou non la subvention ou la contribution. C'est ainsi que si le gouvernement décide, dans le cas susmentionné, que la subvention doit être versée inconditionnellement, celui-ci ne sera pas en mesure d'établir le succès ou l'échec du programme d'aide en se fondant sur des données relatives au groupe visé.

Lorsque le programme en cause se subdivise en petites composantes ou en projets, il pourrait être illlogique de mettre en commun les conclusions de rapports différents. Par surcroît, le genre de questions posées pour mesurer l'efficacité du programme principal peut être totalement différent³¹. Dans le cas, par exemple, d'un programme

qu'elles étaient utiles. Enfin, les gouvernements de tous les pays démocratiques modernes mettent en oeuvre - généralement pour des raisons d'ordre moral - des programmes susceptibles de paraitre inefficaces à beaucoup de contribuables. Citons, comme objet d'un programme de ce genre, le traitement de certains types de troubles mentaux: tout soin prodigué à cet égard est jugé préférable à l'absence complète de soins. L'absence de soins serait perçue comme un aveu d'impuissance par les malades et tendrait à démontrer l'indifférence des spécialistes. Arguer du fait que l'on ne peut pas démontrer raisonnablement l'efficacité du traitement ne constitue peut-être pas, dans de telles conditions, une contribution utile à la résolution du problème. Mais les objectifs généraux du programme sur le plan social peuvent être de montrer que l'on se soucie du sort des malades en cause et que les recherches se poursuivent dans le domaine, de sorte que l'exposé des objectifs pourra être vague.

Après avoir établi que pour certains programmes, surtout de nature sociale, il peut être préférable de ne pas avoir d'objectifs précis, il convient de faire ressortir deux points. Premièrement, pour bien d'autres programmes, les efforts consacrés à la définition et à la mesure des progrès réalisés en fonction des objectifs fixés constitueront un centre d'intérêt fort utile. Omettre l'analyse d'appréciation des objectifs devrait être l'exception et non la règle. Deuxièmement, même si la mise en relief des objectifs n'est pas indiquée dans un cas donné, les autres questions d'évaluation doivent néanmoins être soulevées. Dans de tels cas, en particulier, il devient essentiel de déterminer quels sont les résultats du programme.

6.2 La méthodologie

Parmi les inconvénients courants et les limites importantes qui caractérisent l'évaluation d'un programme, il y a la difficulté que pose l'isolement des effets ou des répercussions du programme par rapport à ceux d'autres programmes ou d'autres facteurs²⁹. En tout temps, notre société est exposée à de nombreuses influences. La plupart de celles-ci ne peuvent être suffisamment contrôlées pour des

En supposant aussi que la "quantification" comme telle ne présente pas de difficultés, l'acceptation du programme peut dépendre de la façon dont différents groupes d'intérêt perçoivent les buts du programme.

Les objectifs mentionnés dans la législation doivent être formulés avec une certaine ambiguïté de sorte qu'elle puisse rallier les suffrages de groupes divers qui croient que la législation sert leur propre cause. Tenir à ce que les buts du gouvernement soient clairement définis peut avoir pour effet de réduire la portée de la législation et de diminuer ainsi la participation du gouvernement dans les programmes sociaux²⁷.

En outre, le fait de mettre l'accent sur la quantification attirera l'attention sur ce qui peut être facilement mesuré. Il est possible que la canalisation subséquente des efforts et des ressources vers ces secteurs se fasse au détriment d'autres besoins plus pressants ailleurs. Ce qui se mesure n'est peut-être pas ce qui présente le plus d'intérêt. Du reste, en mettant l'accent sur les indicateurs mesurables, on risque de faire de ceux-ci les buts des programmes²⁸.

Plusieurs autres raisons peuvent justifier le caractère vague ou imprécis des objectifs d'un programme, indépendamment de l'impossibilité ou de la difficulté foncières à les définir. Ainsi, les objectifs seront parfois flous pour des raisons liées à la sécurité ou à l'unité nationale. Dans d'autres cas, le programme peut être axé sur la prestation d'un service dont on ne sait trop comment l'assurer efficacement, même s'il est sincèrement souhaité. À titre d'exemple, on peut mentionner les programmes destinés à promouvoir la tolérance raciale ou religieuse ou à freiner un comportement jugé indésirable. En pareil cas, les objectifs ne sont pas précisés afin de permettre une mise en oeuvre et une orientation du programme qui soient souples au départ et qu'il y ait place plus tard pour les ajustements nécessaires face aux nouveaux besoins ou aux changements qui surviennent dans le domaine visé par le programme. De cette manière, les ressources peuvent être canalisées vers des activités dont l'expérience a appris

Le problème fondamental, ce n'est pas que certains programmes n'ont pas de buts. Au niveau des politiques, chaque programme a des buts ou des objectifs, si généraux soient-ils. Cependant, il est souvent difficile de fixer des buts précis quantifiables qui permettent à l'évaluateur de relier d'une manière logique les buts des programmes et les objectifs des politiques²⁴. Le Rapport du Comité indépendant de révision, sur les fonctions du Vérificateur général du Canada (Comité Wilson) reconnaissait, en 1975, les grandes difficultés soulevées par l'évaluation des objectifs:

Les problèmes actuels qui se posent quant à l'efficacité des programmes de dépenses gravitent autour de deux grandes questions. Premièrement, comment peut-on définir en termes quantifiables ce que le Parlement entend accomplir en approuvant une nouvelle loi ou en approuvant un nouveau programme par un vote de fonds dans une loi portant affectation de crédits? De nombreux programmes ne se prêtent pas à quantification. Comment peut-on, par exemple, interpréter les objectifs du ministère des Affaires extérieures en termes quantitatifs qui permettent de mesurer objectivement son rendement? Même si certains programmes peuvent être axés sur la réalisation de multiples objectifs, dont certains sont mesurables (par exemple, la réduction du nombre de chômeurs ou l'amélioration du niveau de vie), d'autres ne le sont pas (par exemple, vaincre un sentiment d'injustice). En supposant qu'on puisse résoudre certains de ces problèmes d'identité, il reste toujours la seconde grande question: de quelle façon peut-on mesurer les résultats obtenus? Dans certains cas, l'information statistique actuelle donne des réponses raisonnables mais, dans d'autres, de nouvelles techniques devront être mises au point pour assurer un calcul valable de leurs résultats.

Même lorsque la quantification ne pose pas de problèmes, il peut être très difficile de s'entendre sur les mesures qui correspondent exactement aux buts fixés.

Le système politico-bureaucratique peut faire naître des indicateurs qui minimisent les problèmes ou les rendent plus faciles à résoudre et qui correspondent aux engagements et aux intérêts émotionnels des décideurs et de ceux qui les influencent. Les indicateurs peuvent servir à faciliter l'adoption des politiques souhaitées, à rendre ces politiques le plus attractives possible et à orienter le débat public dans le sens voulu.

l'usage des résultats de l'évaluation: des difficultés surviennent lorsque les gestionnaires et les personnes qui établissent les politiques prennent connaissance des résultats des évaluations et tentent d'appliquer les recommandations qui en découlent.

la valeur du programme: lorsqu'il n'y a pas de conclusion finale claire, cela signifie que le jugement définitif est de nature subjective.

6.1 Les objectifs d'un programme

Autrefois, les évaluations d'efficacité portaient surtout sur l'analyse du degré de réalisation des objectifs d'un programme (tableau 2). Les résultats de ce type d'étude étaient limités. Souvent, l'évaluateur est incapable d'accomplir son travail à cause de la nature floue de nombreux objectifs, et le gestionnaire du programme s'oppose à ce que l'on tente de quantifier et de mesurer les éléments qu'il ou elle croit être essentiellement qualitatifs et non mesurables.

Il est tout simplement impossible de mesurer, et il en sera peut-être toujours de même, les objectifs de certains programmes du secteur public. En Suède, le Bureau national de la vérification a définitivement délaissé l'approche des "objectifs", pour adopter l'évaluation des processus et procédures, en collaboration avec les gestionnaires. Le Vérificateur général de la Suède a fait le point récemment sur cette expérience de 10 ans de la Suède:

Lorsque nous avons commencé à effectuer des vérifications d'efficacité, en Suède, nous attachions énormément d'importance à la question des objectifs. Les vérificateurs devaient se servir des objectifs déterminés par les politiciens comme point de départ pour évaluer dans quelle mesure les activités des organismes gouvernementaux permettaient d'atteindre les dits objectifs. L'expérience nous a convaincus que cette méthode n'est pas efficace. Une activité vise quelquefois la réalisation de plusieurs objectifs et ceux-ci sont souvent confus et mal déterminés. Cette situation n'est pas près de changer. Souvent les politiciens évitent de s'engager en déterminant des objectifs précis et clairs pour leur politiques. Par conséquent, nous sommes abstenus, en règle générale, d'évaluer le degré de réalisation des objectifs... En fait, nous travaillions en nous basant sur l'hypothèse que, plus tard, les vérifications des opérations remplacements les vérifications de l'efficacité en Suède.

Plusieurs gouvernements effectuent des évaluations de leurs programmes. L'importance et la portée de l'évaluation des programmes dépendent des accords constitutionnels historiques d'un pays et de la façon dont les activités gouvernementales ont été dirigées dans son contexte politique. De nombreux pays et de nombreuses institutions de chacun de ces pays ont essayé diverses méthodes d'évaluation avec plus ou moins de succès. Ils ont tous trouvé qu'il est difficile de faire une évaluation de l'efficacité qui puisse être utile. En Suède, plusieurs types d'évaluation se font dans les organismes et par des Commissions royales. Toutefois, le Bureau national de la vérification, centre névralgique de l'évaluation, concentre ses efforts sur les vérifications de système pour déterminer l'économie et l'efficacité, plutôt que sur les évaluations d'efficacité en vue d'améliorer les pratiques de gestion. Le Trésor britannique effectue des évaluations, dans le cadre de ses Program Analysis Reviews, qui servent parfois à la planification, mais elles ne contribuent apparemment pas beaucoup à améliorer les méthodes de gestion. Aux Etats-Unis, le General Accounting Office (GAO) doit procéder à ce qu'il appelle "l'examen des résultats" de tous les programmes. En pratique, il s'avère souvent que les raisons d'être des programmes ne sont pas suffisamment élaborées pour se prêter à un tel examen. Aux Etats-Unis, on fait, toutefois, beaucoup d'évaluations de programme dans les divers ministères et organismes gouvernementaux, tant au niveau fédéral qu'à celui des Etats²¹.

Certains faits se dégagent de l'expérience canadienne et étrangère acquise en matière d'évaluation de programme. Un certain nombre de problèmes ont contribué à refreîner quelque peu les attentes:

- les objectifs du programme: il est difficile de déterminer des objectifs mesurables pour tous les programmes et de créer des mesures valables de réalisation des objectifs.
- la méthodologie: il n'y a pas assez de méthodes standards adéquates d'étude des problèmes sociaux et économiques.
- les contraintes en matière de ressources: toutes les catégories de ressources impliquées dans l'exécution des évaluations sont limitées.

d'études d'évaluation appropriées est évident. En particulier, le client de l'étude et les questions posées déterminent la stratégie d'évaluation à adopter, donc le point central, l'accent et finalement les résultats de l'étude. La conception d'études d'évaluation appropriées nécessite une planification minutieuse avant le début de l'étude. Une étude préparatoire d'évaluation s'impose dans la majorité des cas. Il n'y a pas une stratégie d'évaluation qui soit toujours préférable à une autre. L'utilisation de techniques d'évaluation compliquées ne garantit pas une évaluation plus valable. L'inverse est vrai quelquefois. Il faut effectuer une enquête appropriée, conçue pour répondre au but visé. Ce n'est qu'ainsi que pourront être réduites au minimum les évaluations de programmes inutiles et souvent coûteuses.

Il en découle que le mandat de chaque évaluation doit être rédigé avec soin et englober les questions précises à poser et un résumé des méthodes à utiliser. Ce point est d'une importance capitale si l'on veut s'assurer que les clients de l'étude - les sous-chefs, le Cabinet ou le Parlement - reçoivent des évaluations qu'ils jugent utiles et conformes à leurs besoins.

20

6. LES LIMITES ET PERSPECTIVES REALISTES DES EVALUATIONS DE PROGRAMMES

Les nombreux avantages qu'entraînent des évaluations plus nombreuses et meilleures ont été décrits précédemment dans l'annexe 3. Toutefois, comme le démontre l'expérience canadienne et étrangère, l'exécution et l'utilisation des évaluations sont soumises à d'importantes restrictions. Le présent article expose quelques-unes des principales leçons tirées de cette sphère en plein essor.

5.6 La nécessité d'établir une certaine crédibilité

L'élément le plus important d'une évaluation c'est probablement sa crédibilité. Il faut faire montre d'impartialité, d'exactitude et de compréhension.¹⁸ Le degré de ces qualités varie selon les évaluations. Pour que les résultats soient pris au sérieux, il faut que chaque évaluation inspire confiance au client.

En ce qui a trait aux évaluations faites pour les sous-chefs en application de la politique du Conseil du Trésor du Canada concernant les évaluations, leur crédibilité sera établie à l'aide d'un certain nombre de mécanismes. Les sous-chefs ont un mot important à dire sur l'objet d'une évaluation et sur la manière de l'effectuer, et ils constituent une partie intégrante du processus d'évaluation. Les services qui doivent effectuer les évaluations dans les divers ministères et organismes sont indépendants des services responsables de la gestion des programmes, ce qui leur donne la liberté d'être plus objectifs. Le Bureau du Contrôleur général est l'un des acteurs dont le rôle consiste à s'assurer de l'objectivité et de l'exactitude des évaluations, tâche dont il s'acquitte en expertisant certaines études. Bref, l'examen des évaluations par le public, surtout si la loi sur l'accès à l'information est adoptée, permettra d'accroître la crédibilité de ces dernières aux yeux de la population.

Pour que les gestionnaires de programmes soient satisfaits de l'évaluation de leurs programmes, ils doivent connaître les points forts et les lacunes des techniques d'évaluation utilisées. En outre, il doit leur apparaître évident que les évaluations ont tenu compte des caractéristiques des programmes étudiés.¹⁹

5.7 Résumé

Les exemples figurant dans le tableau 1 et ailleurs dans le texte illustrent la grande variété de questions possibles qui peuvent être étudiées sous la rubrique "évaluation de programmes". Le besoin

Les études d'évaluation pour le compte de clients différents nécessitent des ressources et des calendriers différents et doivent comporter des exigences appropriées quant aux délais. Les évaluations effectuées à l'intention des gestionnaires portent généralement sur les résultats à court terme et doivent être terminées dans les délais prescrits afin que les gestionnaires puissent en tenir compte dans leurs décisions. Autrement, qu'elle que soit la valeur technique de l'étude, si on ne peut l'obtenir au moment voulu, elle devient inutile et sans aucune portée.

Les évaluations destinées au Parlement ou aux organismes centraux doivent également respecter les délais imposés. On peut demander des rapports d'évaluation lorsqu'une question particulière retient l'attention du public et on doit soumettre ces rapports aux clients au moment où ils en auront besoin.¹⁶

Les évaluations destinées aux chercheurs et aux planificateurs exigent souvent plus de travail en raison de l'importance qu'ils accordent aux répercussions des programmes. Dans ces évaluations, on utilise également des techniques analytiques plus complexes et on consulte beaucoup plus de spécialistes.¹⁷

Finalement, les études d'évaluation doivent se justifier quant à leurs coûts. Bien qu'il n'existe aucune règle générale pour déterminer si une étude d'évaluation vaut l'argent qu'on y consacrerait, il faut tâcher ordinairement de déterminer si les avantages qu'on en retirera se rapprocheront des coûts. Une étude d'évaluation ne se justifie aucunement quand on juge clairement que les bénéfices possibles se trouveront inférieurs aux coûts encourus.

5.5 Les ressources et le temps disponibles

Dans l'établissement d'une évaluation appropriée, on doit tenir compte des ressources et du temps disponibles. Ces contraintes peuvent être souvent appréciées lors de la phase de l'étude préparatoire de l'évaluation.

Généralement, l'évaluation peut être conçue en fonction de ressources et de délais d'exécution précis. Il faut quelquefois choisir entre la quantité et la qualité des renseignements que l'on veut obtenir. Par exemple, les résultats d'une évaluation de l'efficacité d'un programme de formation de la main-d'œuvre seraient beaucoup plus précis si on disposait d'un "groupe-témoin", i.e., d'un groupe offrant les mêmes caractéristiques que les participants au programme. Cette méthode entraîne des dépenses plus élevées et prend plus de temps que la technique du "groupe de comparaison", établi dans ce cas-ci d'après les données administratives existantes en utilisant les dossiers rejetés parce que certain(s) critère(s) d'éligibilité n'était(en)t pas satisfait(s). L'équipe chargée de l'évaluation du programme de formation de la main-d'œuvre en 1977 (voir tableau 1), a été confrontée avec un tel choix et elle a opté pour le "groupe de comparaison". Cette méthode a permis d'obtenir des résultats utiles, mais moins probants.

Les études d'évaluation seront plus ou moins coûteuses d'après la profondeur de l'étude effectuée à l'égard de chaque question particulière, le nombre des questions étudiées et la disponibilité de l'information. Dans la plupart des cas, surtout lorsqu'on examine les répercussions des programmes, il faut obtenir des renseignements à l'extérieur du programme. L'obtention de ces renseignements peut être coûteuse et exiger beaucoup de temps, mais il se peut qu'elle soit essentielle à l'utilité de l'étude.¹⁵

par exemple, que l'évaluation de programme faite à l'intention du Parlement, du Cabinet ou d'organismes centraux comporte proportionnellement plus d'études de caractère interministériel que les études faites pour chacun des sous-chefs.

5.4 La méthode d'évaluation

Dans l'établissement d'une stratégie d'évaluation pertinente, le quatrième élément qu'il faut examiner est la méthode ou la technique d'évaluation proposée. La gamme très variée des types d'évaluation et la diversité des clients exigent l'utilisation de méthodes différentes, depuis les analyses quantitatives jusqu'aux recherches qualitatives, selon qu'il s'agit d'évaluations rigoureuses à moins rigoureuses, (voir tableau 1)¹³. Afin de déterminer les résultats d'un programme particulier du gouvernement, on peut très bien utiliser des méthodes analytiques très perfectionnées. Par exemple, s'il s'agit d'une évaluation à l'intention des planificateurs, les techniques de stimulation par ordinateur peuvent être nécessaires pour analyser l'effet des stimulants économiques régionaux sur les provinces Atlantiques. Si l'évaluation est faite à l'intention de la direction d'un programme

On doit reconnaître que pour certains programmes, il peut se révéler impossible d'obtenir des réponses quantitatives satisfaisantes à certaines questions. Cela ne veut pas dire qu'il est impossible d'établir une méthode d'évaluation. On peut toujours utiliser des méthodes quantitatives ou qualitatives plus souples pour une analyse très utile des divers aspects et des résultats d'un programme et pour obtenir au moins des réponses fragmentaires aux questions d'évaluation les plus importantes.

... la méthodologie peut varier; on cherchera alors à établir la pertinence, l'exactitude et le bien-fondé de l'évaluation par rapport à un ou plusieurs éléments de gestion. L'évaluation vise alors principalement les gestionnaires de programmes, qui n'ont pas les mêmes normes que les chercheurs en sciences sociales ou qu'un vérificateur, respectivement, en ce qui concerne la rigueur expérimentale, ou l'établissement d'une preuve.

Compte tenu de la très grande diversité des intérêts et des tâches des clients susmentionnés, lors de l'évaluation d'un programme particulier, il y aura des variantes au niveau de la méthode d'approche adoptée, des questions posées et du genre de résultats obtenus selon le client auquel l'étude est destinée.¹²

Par exemple, en 1977, le Conseil économique du Canada a procédé à l'évaluation des avantages des prestations de la Commission d'assurance-chômage (tableau 1). Le Conseil voulait déterminer l'effet global des modifications apportées au programme et cherchait à mieux comprendre les différents aspects du programme afin d'en améliorer la planification. Si le Parlement évaluait le même programme, il pourrait chercher à en connaître les effets sur les particuliers ou les groupes de prestataires, ou les répercussions dans les petites localités. Un autre client pourrait s'intéresser aux effets du programme à titre de supplément du revenu, et non pas à titre d'assurance, ou encore il pourrait chercher à savoir s'il est possible d'obtenir les mêmes résultats au moyen d'un régime d'impôt négatif. De son côté, un gestionnaire de programme peut s'intéresser à l'amélioration des résultats grâce à des changements graduels dans le programme, au lieu de procéder à des remaniements importants de la politique globale.

Les études d'évaluation des programmes effectuées selon les directives de la Circulaire de politique du Conseil du Trésor 1977-47 (voir l'article 4.2) ont comme client le sous-chef du ministère ou de l'organisme approprié et, s'il est vrai qu'une évaluation exhaustive entraînerait l'examen de toutes les questions formulées dans le tableau 2, les études effectuées traduiraient les intérêts et les préoccupations du sous-chef. Cela apparaîtrait avec évidence dans les questions particulières qui seraient posées et, tout probablement, dans les groupes d'activités cibles de chaque étude. Comme clients des études d'évaluation, le Cabinet ou des organismes centraux pourraient fort bien s'intéresser à des questions différentes et à l'étude de regroupements différents d'activités. Il serait normal,

donnée peut ne pas regrouper les données permettant d'établir la mesure dans laquelle les objectifs ont été atteints si l'étude vise essentiellement à déterminer l'étendue et la nature d'un effet négatif inattendu. Avant chaque étude d'évaluation, il faut déterminer le type d'information que l'on doit recueillir afin de s'assurer que l'étude porte sur les questions pertinentes aux yeux du client. C'est le but principal de l'étude préparatoire d'évaluation faite avant le début de l'évaluation formelle.

Il faut aussi noter que la définition d'un programme aux fins de l'évaluation peut dépendre de la nature de la question d'évaluation qu'on veut soumettre à l'examen. L'examen de la raison d'être d'un programme porterait manifestement sur le programme lui-même alors qu'une évaluation des répercussions du programme sur un groupe client pourrait bien s'étendre à tous les programmes qui intéressent le groupe.

5.3 Le client de l'évaluation

En plus de connaître les questions à poser et le but de l'évaluation, il est indispensable de connaître le client et de comprendre les intérêts en cause. Il peut y avoir également de nombreux utilisateurs de toute étude et les clients de l'étude peuvent vouloir tenir compte des préoccupations des utilisateurs.

Il existe un grand nombre de clients et d'utilisateurs possibles des études d'évaluation de programmes gouvernementaux:

les comités du Cabinet
les média d'information
le Parlement
les planificateurs de programme
le public
les gestionnaires de programme
les chercheurs
les hauts fonctionnaires

meilleure qualité. Le programme serait alors plus efficace. Les solutions de rechange pourraient englober, selon les circonstances, d'autres moyens d'appliquer le programme ou d'autres programmes permettant d'atteindre les objectifs ainsi que les répercussions et les effets intentionnels. On peut aborder ces questions en examinant la possibilité d'économies d'échelle et les effets de la suppression du programme.

Une partie ou la totalité de ces diverses questions peuvent constituer le thème d'une étude d'évaluation particulière. Il faut répondre à toutes ces questions lorsqu'on veut réaliser une étude d'évaluation d'un programme qui soit complète. Evidemment, il n'est pas toujours souhaitable ni même possible de fournir des réponses en profondeur à chacune des questions d'évaluation. Par exemple, l'évaluation des programmes gouvernementaux réalisée en vertu de la politique du Conseil du Trésor du Canada exige que chaque question soit examinée sérieusement et que soit conçue une étude d'évaluation appropriée tenant compte des besoins du client - le sous-chef - de l'ordre prioritaire des questions faisant l'objet de l'évaluation, des avantages prévus dans chaque cas ainsi que des ressources disponibles. Par contre, les études d'évaluation réclamées par le Cabinet ou le Parlement sont susceptibles de porter uniquement sur les questions particulières qui les intéressent.

Les réponses aux questions du tableau 2 montrent quel est le genre d'information à obtenir dans une étude d'évaluation. C'est pourquoi le mandat d'une étude particulière doit être aussi précis que possible quant à ce que l'on désire exactement; il ne faut pas s'en tenir à des énoncés généraux tels que donne le tableau. Par exemple, les buts d'un programme fournissant un service doivent être précis quant au niveau de service visé - quant à la quantité, à la qualité, à la ponctualité, à la fréquence et à l'équité du service¹¹. Bien sûr, d'après la conception de l'étude, il n'est pas nécessaire de recueillir des renseignements à l'égard de chacune des questions: une étude

Tableau 2

ASPECTS FONDAMENTAUX DE L'EVALUATION DE PROGRAMME

Aspects de l'évaluation	Questions fondamentales
Raison d'être du programme	<p>• Dans quelle mesure les objectifs et le mandat du programme sont-ils encore pertinents?</p> <p>• Les activités et les extrants du programme sont-ils conformes à son mandat et reliés de façon vraisemblable à la réalisation des objectifs et à l'obtention des répercussions et effets intentionnels?</p>
Répercussions et effets	<p>• Quels effets et répercussions intentionnels ou non, l'application du programme a-t-elle produits?</p> <p>• Dans quelle mesure le programme est-il une cause de complètement, de double emploi, de chevauchement ou de contradiction par rapport à d'autres programmes?</p>
Réalisation des objectifs	<p>• De quelle façon et dans quelle mesure les objectifs du programme ont-ils été réalisés grâce à sa mise en oeuvre?</p>
Solutions de rechange	<p>• Existe-t-il de meilleurs moyens d'atteindre les objectifs et les effets et répercussions intentionnels recherchés?</p> <p>• Existe-t-il d'autres façons plus rentables de mettre en oeuvre le programme actuel?</p>

L'examen de la raison d'être vise à déterminer si le programme demeure pertinent dans les conditions sociales et économiques du moment et compte tenu de la politique du gouvernement. On remet ici en question l'existence même du programme en se demandant si le programme demeure nécessaire, dans l'hypothèse même où les réalisations du programme sont celles que l'on attend. Une comparaison des activités courantes du programme avec les activités prévues dans le mandat et un examen de la qualité des liens entre les extrants du programme et ses objectifs de même que ses répercussions et effets intentionnels aideront à mieux comprendre la raison d'être du programme.

L'évaluation de l'impact et des répercussions est d'une portée plus générale puisque l'on cherche ici à déterminer le résultat effectif du programme. On s'intéresse aux répercussions immédiates et aux effets à plus long terme du programme, qu'il s'agisse des effets intentionnels et, chose souvent plus intéressante encore, des effets non intentionnels, abstraction faite des objectifs proclamés ou prétendus du programme. Cela comprend, notamment, la répartition des richesses de même que les répercussions et les effets rétrogrades du programme à l'étude et des programmes connexes.

L'analyse de la réalisation des objectifs vise à déterminer, tout d'abord, de quelle façon et dans quelle mesure les objectifs recherchés ont été atteints par suite de la mise en oeuvre du programme. On a souvent considéré à tort que cette question était la seule à intéresser l'évaluation de l'efficacité.

Enfin, l'appréciation des solutions de rechange est un barème pour l'évaluation de la valeur relative du programme. Les objectifs ont pu être atteints, il n'y a peut-être pas de répercussions ou d'effets de nature négative, mais il faut étudier les possibilités soit d'obtenir les mêmes résultats à un moindre coût, soit d'augmenter les coûts avec une production accrue plus que proportionnellement ou de

Après avoir identifié les activités qui vont constituer le programme à évaluer, il est nécessaire de comprendre un peu comment le programme est censé fonctionner et de comprendre sa raison d'être fondamentale. Cela veut dire, comment les activités du programme produisent des résultats de programme, lesquels, à leur tour, se traduisent par des répercussions et des effets particuliers sur la société. Les objectifs du programme peuvent être déterminés en termes de résultats du programme, les répercussions ou effets intentionnels du programme, ou les deux, selon la nature du programme ou le niveau de l'objectif étant définis.

Il faut aussi considérer la maturité du programme. Il est souvent difficile d'évaluer un programme qui vient d'être mis sur pied ou qui n'a pas encore été suffisamment rodé⁹. De plus, le type de services ou de biens que fournit le programme est aussi un facteur que l'évaluateur devra considérer¹⁰, tout comme il devra examiner le genre d'intervention utilisé: dépenses, réglementation ou dépenses fiscales.

5.2 Le genre de questions à poser

Dans l'évaluation d'un programme, on peut soulever une foule de questions et rassembler et analyser une vaste gamme de renseignements relatifs aux activités et aux résultats de celui-ci. (Selon le tableau 1, ces questions peuvent avoir un caractère général et toucher notamment les répercussions du programme ou avoir un caractère plus limité et porter, par exemple, sur l'acquisition de ressources.) Il est indispensable, pour élaborer une étude convenable, de bien saisir les questions précises qu'il faut poser. Cela aide à déterminer l'objet précis et l'orientation de l'étude.

On trouvera au tableau 2 sept questions fondamentales en matière d'évaluation d'un programme basées sur les quatre aspects généraux d'évaluation mentionnés dans l'article 2. En fait, ce tableau donne une définition opérationnelle de l'expression évaluation de programmes employée dans l'administration publique fédérale.

Pour que l'évaluation d'un programme donne soit appropriée, il faut tenir compte:

- de la nature du programme à évaluer;
- du genre de questions à poser;
- du client de l'évaluation;
- de la méthode d'évaluation à adopter;
- des ressources et du temps dont on dispose pour l'évaluation; et
- du besoin d'assurer la crédibilité.

Ces six facteurs aideront les responsables de l'évaluation ainsi que le client à élaborer une stratégie d'évaluation appropriée. Ils sont présentés dans l'ordre, ci-dessous.

5.1 La nature du programme à évaluer

Dans la mise en oeuvre d'une étude d'évaluation, le premier élément dont il faut tenir compte est la définition ou la description de ce qui doit être évalué. Il peut parfois être difficile de déterminer ce qui constitue un "programme" aux fins d'évaluation⁸. Un programme peut ne pas coïncider avec la structure d'organisation du gouvernement ou d'un ministère. Entre autres, un programme à évaluer peut ne pas correspondre avec un programme tel que décrit dans les Prévisions budgétaires annuelles du gouvernement. L'administration publique fédérale a fait un travail considérable pour élaborer des éléments de programmes pertinents aux fins d'évaluation conformément à la politique d'évaluation du Conseil du Trésor du Canada. Ces composantes de programme regroupent des activités ayant un objectif ou un ensemble d'objectifs communs établis au palier du sous-chef, et dont la taille ou les répercussions sont suffisantes pour en faire un élément déterminant de la prise de décision concernant la programmation à l'échelle du ministère sans pour autant se situer à un niveau trop agrégé. (Dans le présent document, on a adopté le terme programme, et non l'expression élément de programme, pour définir ce qui est évalué.) Souvent, la meilleure subdivision en composantes de programmes aux fins d'évaluation n'est évidente qu'après une analyse détaillée des activités du ministère.

été entreprise conformément à la politique de 1977 du Conseil du Trésor du Canada. Plusieurs de ces études étaient déjà en chantier voire même complétées. Dans de nombreux cas également, les clients de ces évaluations n'étaient pas des sous-chefs. Ces évaluations avaient pour objet de déterminer l'incidence des programmes sur la société ou d'examiner les éléments et les activités qui composent le programme. Il est évident aussi que les techniques d'évaluation varient d'une étude à l'autre et comportent des analyses économétriques ou des entrevues avec les bénéficiaires du programme ou encore avec les spécialistes en la matière.

Le large éventail d'évaluations possibles renforce le besoin de préciser davantage l'expression "évaluation des programmes". Pour toute évaluation, il importe de s'assurer qu'on élabore une stratégie appropriée pour traiter des aspects les plus pertinents et pour satisfaire les besoins du client de l'étude.

5. LA PLANIFICATION D'UNE EVALUATION DE PROGRAMME APPROPRIÉE

La planification d'une étude d'évaluation est peut-être la phase la plus difficile. En plus d'être coûteuses tant sur le plan des ressources financières que sur celui des ressources humaines, l'expérience a prouvé que bon nombre d'études d'évaluation effectuées par le passé sont inutilisables (voir article 6.4). S'il est impossible de les employer, c'est, en partie, parce qu'on n'a pas assez examiné, avant de procéder à une étude d'évaluation, les besoins qui existent, les possibilités offertes et celles qu'on retiendra. Cette planification visant à déterminer le point sur lequel on doit se concentrer et la méthodologie à employer - ce qu'on appelle ici une étude préparatoire d'évaluation - est indispensable à la conception d'une évaluation appropriée.

Tableau 1

QUELQUES ETUDES D'EVALUATION DE PROGRAMMES DU GOUVERNEMENT

Programme évalué	Date	Cliënt	But de l'évaluation	Méthode de l'évaluation
Programme de logement du ministère des Travaux publics	1978	Comité du Sénat sur les finances	Décrire la capacité du programme à rencontrer ses objectifs et d'évaluer les points forts et les points faibles du programme	Interviews d'experts et de fonctionnaires
Programme de subventions à l'encouragement à la recherche et au développement	1976	Ministère de l'Industrie et du Commerce	Déterminer l'efficacité des subventions à la R et D	Analyse économétrique
Ministère des Pêches et de l'Environnement	1977-78	MPE et Conseil du Trésor Canada	Identifier par l'analyse des dépenses les ressources requises pour lancer de nouvelles initiatives	Analyse des activités du ministère
Programme de formation de la main-d'oeuvre du Canada	1978	Emploi et Immigration, Conseil du Trésor Canada et Finances	Examiner les activités du programme dans le contexte du mandat du programme	Etudes par échantillon-nage des participants
Programme de prestations de la Commission d'assurance-chômage	1977	Conseil économique	Déterminer les effets des prestations de l'AC sur la main-d'oeuvre active	Analyse économétrique

A une plus grande échelle, l'évaluation permettra de mieux comprendre les activités qui se rattachent aux programmes ainsi que les véritables répercussions de ces derniers. Cela contribuera à:

- améliorer l'imputabilité en ce qui concerne les programmes du gouvernement.
- renforcer l'aptitude du gouvernement à attribuer des ressources rares, pour en tirer les meilleurs résultats possibles;
- accroître la capacité du gouvernement à prévoir et, de ce fait, à concevoir de meilleurs programmes, pour atteindre les objectifs en matière de politique;
- préciser l'apport des programmes à la mise en oeuvre des politiques;
- améliorer la gestion et l'administration des programmes du gouvernement et, de ce fait, à fournir un meilleur gouvernement aux Canadiens;

Voilà les conséquences directes d'évaluations plus nombreuses et meilleures, en plus de deux autres répercussions d'ordre général:

- la participation régulière à l'évaluation des programmes encouragera le personnel affecté à un programme à examiner d'un oeil critique leurs propres hypothèses et leur comportement et, de ce fait, à trouver et à expérimentier des moyens de mieux s'acquitter de leurs fonctions; et
- l'acceptation, par les particuliers et les groupes concernés, du processus d'évaluation des programmes comme étant un procédé valable et utile pour déterminer la valeur des programmes à l'égard de la société contribuerait à réduire les malentendus et les désaccords quant aux répercussions d'un programme donné.

4. QUELQUES EVALUATIONS RECENTES DES PROGRAMMES DU GOUVERNEMENT

Le tableau 1 montre la nature variée des évaluations portant sur les programmes du gouvernement fédéral. Les études d'évaluation sont faites par des personnes différentes et pour des motifs différents. Par exemple, il est clair qu'aucune des études figurant dans le tableau 1 n'a

Le but de l'évaluation de programme au sein du gouvernement fédéral, poursuivi par cette politique, est d'assurer que les sous-chefs des ministères et des organismes possèdent des renseignements suffisants sur les résultats de leurs programmes afin de pouvoir:

- rendre compte des deniers publics dont ils ont la responsabilité;
- prendre des décisions avisées sur la gestion et les ressources nécessaires aux programmes; et
- fournir des avis de qualité aux ministères.

Des modifications de la structure du système ont découlé de la progression de la fonction de l'évaluation des programmes au sein du gouvernement fédéral. Le Bureau du Contrôleur général a été institué en juin 1978. Ce Bureau a de nombreuses attributions et responsabilités en matière de pratiques et de réglementations administratives dans les domaines de la gestion des finances et des procédures touchant à l'évaluation des programmes du gouvernement du Canada. Le Bureau doit, plus particulièrement, s'assurer que les ministères et organismes ont établi, à des fins d'évaluation, des procédures conformes aux directives du Conseil du Trésor du Canada.

3. LES AVANTAGES RESULTANT DE L'EVALUATION DES PROGRAMMES

De plus en plus, on reconnaît que l'évaluation périodique des programmes constitue la clé d'une bonne gestion des ministères gouvernementaux. On a observé que:

L'évaluation est nécessaire en raison de la rareté des ressources.
...l'évaluation sous-entend la répartition des ressources limitées entre différentes activités.

en réponse aux grandes priorités ou questions en matière de planification de politiques, plutôt que dans le cadre de l'examen permanent des programmes existants, exécuté par les gestionnaires ministériels.

Bien que les gestionnaires du gouvernement aient toujours effectué des études d'évaluation, la première évaluation systématique et périodique de programme remonte à 1977 avec la parution de la circulaire n° 1977-47 du Conseil du Trésor "Évaluation des programmes, effectuée par les ministères et organismes". Cette politique se fonde, dans une certaine mesure, sur les meilleures méthodes déjà adoptées dans de nombreux ministères et organismes. Voici l'énoncé principal de la politique:

Les ministères et organismes fédéraux devront périodiquement étudier leurs programmes afin d'évaluer l'efficacité avec laquelle les objectifs sont atteints et l'efficacité avec laquelle les programmes sont dirigés.

La politique se focalise sur l'évaluation des programmes par opposition, en particulier, à l'évaluation des systèmes et des procédures qui sont, généralement, examinés lors de la vérification interne.⁵ En vertu de cette politique, l'évaluation de programme est censée porter sur un nombre de questions fondamentales:

- La raison d'être du programme:
Le programme est-il encore pertinent?
- Répercussions et effets:
Qu'est-il arrivé comme résultat du programme?
- Réalisation des objectifs:
Le programme a-t-il donné les résultats escomptés?
- Solutions de rechange:
Y a-t-il de meilleurs moyens d'atteindre les mêmes résultats?

des programmes. Les rapports de la Commission, publiés en 1962 et 1963, ont donné lieu à une décentralisation de la gestion et du contrôle financiers et à l'élaboration de méthodes de gestion plus modernes. En plus de la volonté habituelle de faire une vérification, on cherchait à savoir si un programme respectait l'esprit de la loi et ses objectifs².

Une nouvelle approche en matière de gestion financière, a été instituée en 1968. Le système de rationalisation des choix budgétaires (RCB, en anglais PBS), devait aider les gestionnaires ministériels à exercer leurs nouvelles fonctions, assumer de plus grandes responsabilités relativement à la gestion de leurs affaires et donner un meilleur aperçu des objectifs fondamentaux des ministères. "En apportant beaucoup de soin à définir les objectifs des programmes et à établir les coûts totaux de chaque programme"³, la RCB devait introduire la planification et la programmation dans le processus budgétaire. Il devait fournir au Conseil du Trésor du Canada un cadre facilitant l'allocation des ressources grâce à sa base pluriannuelle⁴. En pratique, on a éprouvé des difficultés en essayant de rendre le système opérationnel et la RCB ne s'est pas montrée à la hauteur des espérances qu'elle avait fait naître.

Toutefois, la RCB a effectivement contribué à mettre l'accent sur l'examen des résultats visés par les programmes. Depuis 1970-1971, ce désir a été traduit dans le Budget des dépenses du gouvernement fédéral qui a établi un classement recoupé des dépenses selon l'organisme, le programme et l'article de dépenses.

Les années 70 ont été témoin de l'essor de l'évaluation partout au gouvernement avec les prévisions et les études de programmes, la vérification opérationnelle et les évaluations quantitatives détaillées et structurées de l'efficacité de certains programmes. Ces dernières évaluations structurées ont été effectuées d'une façon hautement sélective,

2. L'EVOLUTION DE L'EVALUATION DES PROGRAMMES AU SEIN DU GOUVERNEMENT FEDERAL

L'étude et l'évaluation des programmes en place ont toujours fait partie du rôle de la gestion gouvernementale. La nature et le contenu des études et des évaluations ont évolué. Auparavant l'intérêt portait principalement sur les ressources affectées à un programme, soit les sommes dépensées et les ressources humaines employées. Maintenant, l'évaluation porte sur la manière dont les ressources sont utilisées, sur les buts des programmes, et sur leurs résultats et leurs répercussions sur la société.

Ce déplacement de l'intérêt s'avère nécessaire. On a fait appel au gouvernement au cours des trois dernières décennies pour qu'il offre de plus en plus de biens et services au public. Bon nombre de programmes gouvernementaux actuels rendent compte d'efforts importants et structures, déployés afin d'améliorer les conditions de vie dans une société de plus en plus complexe. A mesure que les activités gouvernementales ont augmenté, les objectifs et les répercussions de tels programmes ont fait l'objet d'examen publics plus minutieux.

L'inquiétude croissante qu'ont soulevée les objectifs et par conséquent les résultats des dépenses du gouvernement, remonte à la Commission royale d'enquête sur l'organisation du gouvernement (Commission Glassco) en 1960. Avant la Commission Glassco, l'analyse des programmes reposait principalement sur la vérification annuelle des dépenses gouvernementales. Le Vérificateur général réalisait une post-vérification pour le Parlement, tandis que le Contrôleur du Trésor procédait à une vérification préliminaire. Ces vérifications ont permis de s'assurer de la probité des dépenses (les intrants du programme) mais elles n'étaient pas axées sur les buts et les résultats

1. QU'EST-CE QU'UNE EVALUATION DE PROGRAMME?

Il est impossible de dégager un consensus sur ce que doit comporter une évaluation de programme. La terminologie utilisée dans les publications spécialisées est extrêmement variée et constitue la source de nombreux malentendus. En termes simples, l'expression évaluation de programme désigne l'examen systématique des programmes gouvernementaux en place pour en déterminer la pertinence dans le contexte présent et l'efficacité. Cela comporte la collecte systématique de toute information dont on peut établir l'exactitude ainsi que la démonstration évidente des résultats d'un programme. Le Bureau du Contrôleur général a développé la position suivante quant à la signification qu'il entend donner à l'évaluation des programmes du gouvernement:

L'évaluation des programmes est l'examen périodique, indépendant et objectif d'un programme, en vue de déterminer l'à-propos des objectifs, de la conception et des résultats, intentionnels ou non, de ces programmes à la lumière des circonstances de l'heure. L'évaluation remet en cause l'existence même du programme. La raison d'être du programme, ses répercussions et son coût-efficacité par rapport à d'autres solutions de rechange sont autant d'aspects fondamentaux de l'évaluation.

Dans ce contexte, un programme est un groupe d'activités, de pouvoirs ou de fonctions du gouvernement, qui peut être rattaché d'une manière logique à un objectif ou à un ensemble donné d'objectifs. Une étude d'évaluation de programme est un examen documenté de certains aspects choisis d'un programme et de son efficacité et un rapport d'évaluation est un document de l'étude et de ses conclusions. L'évaluation des programmes est la fonction de gestion qui consiste à évaluer un programme relativement à toutes les questions pertinentes (voir le tableau 2). Le présent article s'intéresse surtout aux études d'évaluation plutôt qu'à la fonction d'évaluation de programme. Des explications de ces définitions sont fournies plus loin (article 5), après la présentation de plusieurs nouveaux concepts.

L'ÉVALUATION DES PROGRAMMES: EXAMEN DE LA QUESTION

Le présent document a pour but d'établir un terrain d'entente et un cadre de référence permettant d'étudier la question de l'évaluation des programmes. Une version antérieure de ce document a servi de document de base relativement à l'examen de plusieurs initiatives en matière d'évaluation. Cette documentation peut servir aussi d'introduction à deux documents en préparation au Bureau du Contrôleur général, qui sont intitulés respectivement Guide sur la fonction de l'évaluation de programme dans les ministères et organismes fédéraux et Principes de l'évaluation de programme par les ministères et organismes fédéraux.

Ce document de travail commence par exposer en quoi consiste l'évaluation de programmes (article 1); il énonce ensuite l'évolution de l'évaluation dans l'administration publique fédérale (article 2), en définit les avantages (article 3), donne des exemples d'évaluations récentes de programmes du gouvernement (article 4), examine les principaux éléments dont il faut tenir compte lorsqu'il s'agit de planifier une évaluation satisfaisante (article 5), et examine les limites de la réalisation et de l'utilisation des évaluations (article 6). Le document fournit en conclusion (article 7) une liste des caractéristiques d'évaluations menées à bien.

Ce document traite des sujets susmentionnés, la plupart du temps, en termes généraux, et il ne se limite pas à décrire l'évaluation des programmes telle qu'elle est effectuée au sein de l'administration publique fédérale. Dans les cas appropriés, il fait état des méthodes particulières d'évaluation de programmes, par exemple, lorsqu'il est question de la politique du Conseil du Trésor du Canada concernant l'évaluation des programmes ou les études d'évaluation qui pourraient être faites pour les organismes centraux, le Cabinet, ou le Parlement. Les renvois de la présente partie donnent des indications et des renseignements supplémentaires sur cette question difficile et complexe.

L'EVALUATION DES PROGRAMMES: EXAMEN DE LA QUESTION

Table des matières

Page

1.	QU'EST-CE QU'UNE EVALUATION DE PROGRAMME?	2
2.	L'EVOLUTION DE L'EVALUATION DES PROGRAMMES AU SEIN DU GOUVERNEMENT FEDERAL	3
3.	LES AVANTAGES RESULTANT DE L'EVALUATION DES PROGRAMMES	6
4.	QUELQUES EVALUATIONS RECENTES DES PROGRAMMES DU GOUVERNEMENT	7
5.	LA PLANIFICATION D'UNE EVALUATION DE PROGRAMME APPROPRIEE	9
6.	LES LIMITES ET PERSPECTIVES REALISTES DES EVALUATIONS DE PROGRAMME	21
7.	QU'EST-CE QUI CONTRIBUE AU SUCCES D'UNE EVALUATION	33
	NOTES	35
	6.1 Les objectifs d'un programme	23
	6.2 La méthodologie	26
	6.3 Les restrictions sur le plan des ressources	28
	6.4 L'utilisation des résultats des évaluations	29
	6.5 L'établissement de la valeur d'un programme	32
	5.1 La nature du programme à évaluer	10
	5.2 Les genres de questions à poser	11
	5.3 Le client de l'évaluation	15
	5.4 La méthode d'évaluation	17
	5.5 Les ressources et le temps disponibles	18
	5.6 La nécessité d'établir une certaine crédibilité	20
	5.7 Résumé	20

Direction de l'évaluation des
programmes
Bureau du Contrôleur général
du Canada
Février 1981

L'EVALUATION DES PROGRAMMES: EXAMEN DE LA QUESTION

Document de travail
de la direction 81-001

L'ÉVALUATION DES PROGRAMMES EXAMEN DE LA QUESTION

DOCUMENT DE TRAVAIL